Plant Genetics and Genomics: Crops and Models 11

Andrew H. Paterson *Editor*

Genomics of the Saccharinae

Springer

Plant Genetics and Genomics: Crops and Models

Volume 11

Series Editor
Richard A. Jorgensen

For further volumes:
http://www.springer.com/series/7397

Andrew H. Paterson
Editor

Genomics of the Saccharinae

Springer

Editor
Andrew H. Paterson, Ph.D.
Plant Genome Mapping Laboratory
and Departments of Crop and Soil Science
Plant Biology, and Genetics
University of Georgia
Athens, GA, USA

ISBN 978-1-4419-5946-1 ISBN 978-1-4419-5947-8 (eBook)
DOI 10.1007/978-1-4419-5947-8
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2012942104

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

This volume is dedicated to the memories of James E. Irvine and Keith F. Schertz, Saccharinae scientists who inspired and mentored the editor, many of the chapter authors, and many others.

Preface

Among flowering plants, several "warm-season" grasses are the most efficient at fixing atmospheric carbon, thanks to "C4" photosynthesis, a complex combination of biochemical and morphological specializations discovered in sugarcane that increase net carbon assimilation at high temperature. The Saccharinae clade of grasses is of singular importance, including one cereal that is fifth in importance among the world's grain crops, as well as three leading biofuel crops, and several of the world's most noxious weeds.

Sorghum bicolor L. Moench. is a leading cereal, fifth in importance among the world's grain crops. Introduced into the USA about 200 years ago, sorghum is grown on 8–10 million acres and has a farm-gate value of ~$1 billion/yr. *Sorghum* is unusually tolerant of drought, a feature essential in the US Southern Plains that often receive too little rain for other grains. In arid countries of northeast Africa, sorghum contributes 26–39% of calories in the human diet. Increased demand for limited fresh water, along with rising global temperature and aridity, suggest that sorghum will be of growing importance.

Expansion of agriculture to provide plant biomass for production of fuels and/or feedstocks will require additions to our present repertoire of crops. The Saccharinae clade of grasses shows singular promise, including three leading biofuel crops, *Saccharum* (sugarcane, the worlds #1 fuel ethanol crop), *Sorghum* (currently the #2 source of seed-based fuel ethanol in the USA, and a promising potential source of cellulosic ethanol), and *Miscanthus*, a promising potential cellulosic ethanol crop with much higher yield than another leading candidate, switchgrass, in the US Midwest. Its adaptability to continental Europe shows the feasibility of producing *Miscanthus* in temperate latitudes.

The *Sorghum* genus also includes one of the world's worst weeds "Johnsongrass" (*S. halepense*), a naturally occurring polyploid hybrid that reduces yields of many crops by up to 45 %. The first federal appropriation for weed control research targetted Johnsongrass. Functional genomic data may lead to new strategies for environmentally benign plant growth regulation, suppressing weed dispersal. Better understanding of reproductive barriers in sorghum may lead to strategies to reduce risk that transgenic *S. bicolor* outcrosses with *S. halepense*.

An important breeding line of *Sorghum bicolor* recently became only the second monocot to have its genome essentially fully sequenced, providing an important complement to the previously sequenced genome of rice and opening new doors into the study and improvement of members of the clade. As a model for the tropical grasses, sorghum is a logical complement to *Oryza* (rice). Sorghum is representative of warm-season grasses in that it has "C4" photosynthesis, while rice is more representative of temperate grasses, using "C3" photosynthesis. The ~740 megabase sorghum genome, with ~90 % of DNA and ~98 % of genes placed in their chromosomal context, is a logical bridge to the ~2,500 megabase maize genome that is also being sequenced, and the ~4,000 megabase genome of sugarcane, the world's leading biomass/biofuels crop. Sorghum shared common ancestry with maize (12–15 million years ago, mya) and sugarcane (5–9 mya) much more recently than rice (42–47 mya). The most recent genome duplication in sorghum appears to be ~70 mya versus ~12 mya in maize and <5–9 mya in sugarcane with lower genetic redundancy promising a higher success rate in relating sorghum genes to phenotypes.

For a multitude of reasons—invigorated interest in biofuels, concerns about a looming worldwide water crisis, the need for more precise and more environmentally benign methods of weed control—the Saccharinae clade has seen a resurgence of interest in the past few years. The Saccharinae have an important role to play in a more bio-based economy and a more sustainable agroecosystem. Sequencing of additional members of the clade has begun, building on their rich histories of conventional breeding and genetics research, but constrained by the challenges of their large and complex genomes. In this book, we seek to share with you, the reader, our enthusiasm about the advances in genetics and genomics of the Saccharinae of the past few years and those that loom on the horizon.

In closing, a clarification of nomenclature is important. As described in detail in Chap. 1, the taxonomic nomenclature of the species that are the focus of this book remains unclear. For the purposes of this book, the authors have been encouraged to adopt the view expressed and explained in Chap. 1, that it appears appropriate that subtribe Sorghine (presently including sorghum) should be merged into subtribe Saccharinae (including Saccharum and Miscanthus). Accordingly, the chapter authors have been encouraged to refer to the Saccharinae as inclusive of all three taxa.

Athens, GA, USA Andrew H. Paterson, Ph.D.

Contents

Contributors

Getu Beyene Texas AgriLife Research, Texas A&M System, College Station, Weslaco, TX, USA

International Laboratory for Tropical Agricultural Biotechnology (ILTAB), Donald Danforth Plant Science Center, 975 North Warson Road, St. Louis, MO, USA

Patrick J. Brown Department of Crop Sciences, University of Illinois, Urbana, IL, USA

Marco T. Buenrostro-Nava Texas AgriLife Research, Texas A&M System, College Station, TX, USA

Surinder Chopra Crop and Soil Science Department, Pennsylvania State University, University Park, PA, USA

Tom Clemente Department of Agronomy & Horticulture, Center for Plant Science Innovation, Center for Biotechnology, University of Nebraska–Lincoln, Lincoln, NE, USA

Ian S. Curtis Texas AgriLife Research, Texas A&M System, College Station, TX, USA

Jeff A. Dahlberg U.S. Sorghum Checkoff Program, Lubbock, TX, USA

Mona B. Damaj Texas AgriLife Research, Texas A&M System, College Station, TX, USA

Mukesh K. Dhillon Division of Entomology, Indian Agricultural Research Institute (IARI), New Delhi, India

Ismail Dweikat Department of Agronomy & Horticulture, University of Nebraska–Lincoln, Lincoln, NE, USA

Dean Engler Mendel Biotechnology, Inc, Hayward, CA, USA

F. Alex Feltus Department of Genetics and Biochemistry, Clemson University, Clemson, SC, USA

Savio Siqueira Ferreira Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil

Brhane Gebrekidan Agriculture Working Group, Ethiopian Academy of Sciences, Addis Ababa, Ethiopia

Alan R. Gingle Plant Genome Mapping Laboratory, University of Georgia, Athens, GA, USA

Martha T. Hamblin Institute for Genomic Diversity, Cornell University, Ithaca, NY, USA

Jagger Harvey Biosciences eastern and central Africa-International Livestock Research Institute (BecA-ILRI) Hub, Nairobi, Kenya

Arlene Howe Center for Biotechnology, University of Nebraska–Lincoln, Lincoln, NE, USA

Monsanto Company, Chesterfield, MO, USA

Yinghua Huang Plant Science Research Laboratory, USDA–ARS, Stillwater, OK, USA

Katrin Jakob Mendel Biotechnology, Inc, Hayward, CA, USA

Russell W. Jessup Department of Soil & Crop Sciences, Texas A&M University, College Station, TX, USA

John A. Juvik Department of Crop Sciences, University of Illinois, Urbana, IL, USA

Segenet Kelemu Biosciences eastern and central Africa-International Livestock Research Institute (BecA-ILRI) Hub, Nairobi, Kenya

Elizabeth A. Kellogg Department of Biology, University of Missouri-St. Louis, St. Louis, MO, USA

Changsoo Kim Plant Genome Mapping Laboratory and Departments of Crop and Soil Science, Plant Biology, and Genetics, University of Georgia, Athens, GA, USA

Clarissa T. Kimber Department of Geography, Texas A&M University, College Station, TX, USA

Stephen Kresovich Department of Biology, University of South Carolina, Columbia, SC, USA

Tejinder Kumar Department of Agronomy & Horticulture, University of Nebraska–Lincoln, Lincoln, NE, USA

Qi Lin Institute of Botany, Chinese Academy of Sciences, Beijing, China

Clint W. Magill Department of Plant Pathology and Microbiology, Texas A&M University, College Station, TX, USA

T. Erik Mirkov Texas AgriLife Research, Texas A&M System, College Station, TX, USA

Paul H. Moore Hawaii Agriculture Research Center, Kunia, HI, USA

Seth C. Murray Department of Soil and Crop Sciences, Texas A&M University, College Station, TX, USA

Bongani Kaiser Ndimba Proteomics Research Group, Department of Biotechnology, University of the Western Cape, Bellville, South Africa

Proteomics Unit, Biotechnology Platform, Agricultural Research Council, Stellenbosch, South Africa

Rudo Ngara Proteomics Research Group, Department of Biotechnology, University of the Western Cape, Bellville, South Africa

Milton Yutaka Nishiyama-Jr Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil

Andrew H. Paterson Plant Genome Mapping Laboratory and Departments of Crop and Soil Science, Plant Biology, and Genetics, University of Georgia, Athens, GA, USA

Ana Saballos Genetics Institute and Agronomy Department, University of Florida, Gainesville, FL, USA

Erik J. Sacks Department of Crop Sciences, University of Illinois, Urbana, IL, USA

Paloma Mieko Sato Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil

Shirley Sato Center for Biotechnology, University of Nebraska-Lincoln, Lincoln, NE, USA

Hari C. Sharma International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, Andhra Pradesh, India

Glaucia Mendes Souza Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil

J. Ryan Stewart Department of Crop Sciences, University of Illinois, Urbana, IL, USA

Haibao Tang Plant Genome Mapping Laboratory and Departments of Crop and Soil Science, Plant Biology, and Genetics, University of Georgia, Athens, GA, USA

Tom L. Tew USDA, ARS, Houma, LA, USA

Mitchell R. Tuinstra Department of Agronomy, Purdue University, Richmond, IN, USA

Wilfred Vermerris Genetics Institute and Agronomy Department, University of Florida, Gainesville, FL, USA

Department of Agricultural & Biological Engineering and Laboratory of Renewable Resources Engineering, Purdue University, West Lafayette, IN, USA

Fabio Vicente Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil

Ming-Li Wang PGRCU, USDA-ARS, Griffin, GA, USA

Pohao Wang Crop and Soil Science Department, Pennsylvania State University, University Park, PA, USA

Xiyin Wang Plant Genome Mapping Laboratory and Departments of Crop and Soil Science, Plant Biology, and Genetics, University of Georgia, Athens, GA, USA

Center for Genomics and Computational Biology, School of Sciences, Hebei Polytechnic University, Tangshan, Hebei, China

Zhanguo Xin Plant Stress and Germplasm Development Unit, USDA-ARS, Lubbock, TX, USA

Toshihiko Yamada Field Science Center for Northern Biosphere, Hokkaido University, Hokkaidō, Japan

Jianming Yu Department of Agronomy, Kansas State University, Manhattan, KS, USA

Part I
Natural History and Genetic Diversity

Chapter 1
Phylogenetic Relationships of Saccharinae and Sorghinae

Elizabeth A. Kellogg

Abstract Multiple taxonomic and phylogenetic studies have been conducted on sugarcane, *Miscanthus*, and sorghum, but to date the results have been contradictory and somewhat confusing. A few generalities have emerged. The Andropogoneae is clearly monophyletic. *Saccharum* and *Miscanthus* are closely related to each other. Their relationship with *Sorghum* is less clear, although they are probably more closely related to *Sorghum* than any of them is to maize or to *Andropogon* and its immediate relatives. The phylogeny of Andropogoneae is largely unresolved, which leads to a number of problems of taxonomic nomenclature. The solution will require considerably more phylogenetic data on a much broader set of species than has been sampled to date.

Keywords Evolution • Phylogeny • Andropogoneae • *Miscanthus* • Polyploid • Classification • Sorghum • Sarga

1 Introduction

Sugarcane, *Miscanthus*, and sorghum are all members of the grass family, Poaceae. The first two are currently placed in the subtribe Saccharinae, whereas sorghum is often given its own subtribe, Sorghinae; both subtribes belong to the tribe Andropogoneae. Understanding the relationships of the three taxa is complicated by considerable phylogenetic uncertainty and taxonomic confusion. As described below, the phylogeny represents a difficult phylogenetic problem, in which rapid speciation early in the history of the group led to a phylogenetic tree with many very short internal branches. Disentangling this history has been complicated by a

E.A. Kellogg (✉)
Department of Biology, University of Missouri, One University Blvd, St. Louis, MO 63121, USA
e-mail: tkellogg@umsl.edu

A.H. Paterson (ed.), *Genomics of the Saccharinae*, Plant Genetics and Genomics: Crops and Models 11, DOI 10.1007/978-1-4419-5947-8_1,

limited sample of species in all studies to date. In addition, the group includes many polyploids, sugarcane and *Miscanthus* among them, whose evolution is likely to be highly reticulate; while molecular phylogenetic approaches can disentangle net-like histories, this has not yet been attempted in the Andropogoneae.

In this chapter, I take a hierarchical approach, beginning with the placement of Saccharinae and Sorghinae in the angiosperms, and noting the characteristics that have originated at various stages in evolution. I then consider the placement of the groups within the phylogeny of Andropogoneae, and discuss some of the implications and limitations of the phylogeny for understanding morphological and chromosomal evolution. In the final section, I consider the implications of the phylogenetic data for classification. Although I may not eliminate any taxonomic confusion, I hope to explain at least why it has occurred and what information might address the problems in the future. One likely conclusion from the available data is that the subtribe Saccharinae will ultimately be merged with the subtribe Sorghinae, and that the former name will take precedence.

2 Placement of Sugarcane, *Miscanthus*, and Sorghum within the Angiosperms

The grasses are angiosperms, and are members of the large monocot clade, which includes about 20 % of known flowering plants. Within the monocots, the grasses belong to the commelinid clade (Fig. 1.1). Thus sugarcane, *Miscanthus*, and sorghum all inherit the molecular and morphological characteristics of each of the larger groups to which they belong (summarized in Stevens 2008). As angiosperms, they have ovules enclosed in ovaries, double fertilization, and the many familiar angiosperm characteristics. As monocots they have a single cotyledon, and vascular bundles scattered in the stem. As commelinid monocots, they have cell walls that contain ferulic acid such that they fluoresce under ultraviolet light. Most commelinid monocots, including sugarcane, *Miscanthus*, sorghum, and the grasses, also produce silica bodies (SiO_2) in their leaves and have bracteate inflorescences. The stomata of the commelinids are characteristically paracytic, meaning that they have subsidiary cells that are parallel to the long axis of the stomate. (Some species have tetracytic stomata, which have two additional subsidiaries, one at either end with the long axis of the cell perpendicular to the opening of the stomate.) The commelinids, except for the palms, accumulate starch in the endosperm; the endosperm that is the source of much human nutrition is thus an ancient characteristic that has been retained in the grasses.

Within the commelinids, sugarcane, *Miscanthus* and sorghum are members of the order Poales (Fig. 1.1). This is a monophyletic group of 17 families, all of which have silica bodies in their epidermis. All members of this group also are characterized by having nuclear endosperm, in which multiple nuclear divisions occur before cell walls are formed. A clade within the Poales includes the grass family (Poaceae), plus the families Anarthriaceae, Centrolepidaceae, Restionaceae, Flagellariaceae, Ecdeiocoleaceae, and Joinvilleaceae; together these are known as the graminoid Poales (Campbell and Kellogg 1987; Kellogg and Linder 1995) (Fig. 1.2). As members of

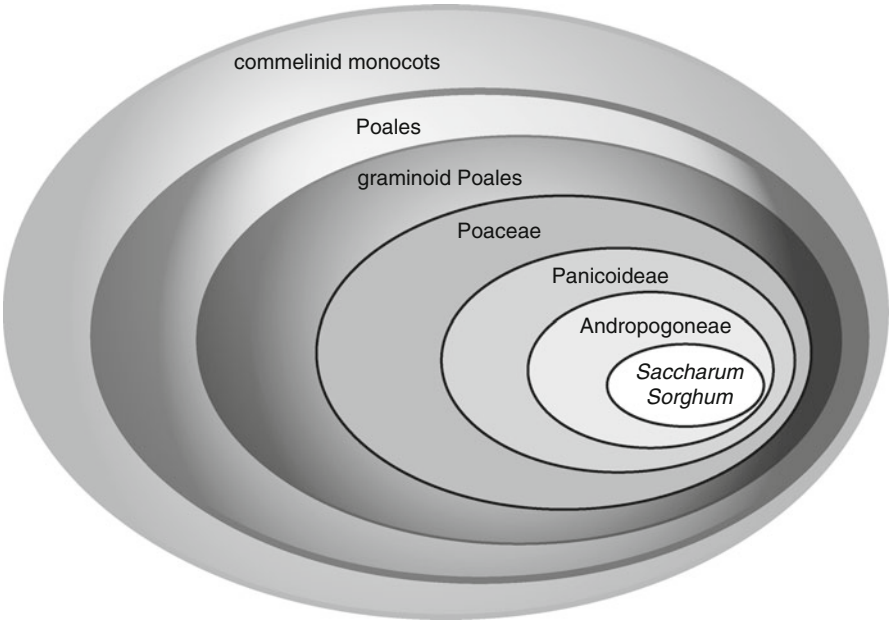

Fig. 1.1 Venn diagram showing the nested relationships of the commelinid clade, Poales, and included taxa

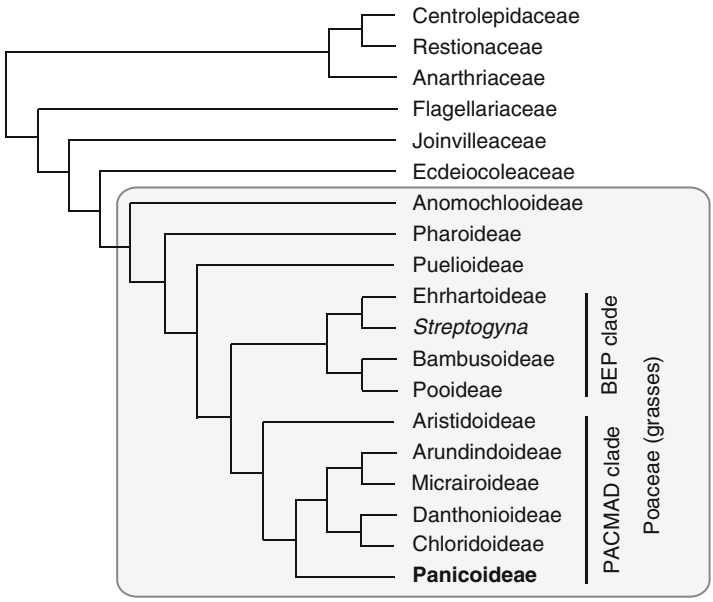

Fig. 1.2 Phylogeny of the graminoid Poales, including relationships among the subfamilies of grasses, following Christin et al. (2008)

the graminoid Poales, the Saccharinae and Sorghinae have distichous leaves, primary cell walls with (1-3,1-4)-ß-D-glucans, and sieve tube plastids with cuneate crystals; the pollen has a single pore (monoporate) surrounded by a raised ring (annulus). All graminoid Poales have one anatropous ovule per carpel; the number of locules is often reduced to one. Most members of the clade are wind pollinated, and thus have tiny flowers with few or no tepals, and feathery stigmas.

As members of the family Poaceae, sugarcane, *Miscanthus*, and sorghum have a grass embryo, which is far more differentiated than embryos in other monocots, and is the result of a heterochronic shift in the timing of embryo maturation relative to the maturation of the seed (Kellogg 2000b). The embryo has a clear apical meristem and several seedling leaves, as well as a unique haustorial organ, the scutellum, which is thought to be a highly modified cotyledon. Like almost all grasses, sugarcane, *Miscanthus*, and sorghum have flowers in spikelets, clusters of one or more flowers with the whole cluster subtended by two bracts, the glumes.

Characteristics of morphology plus extensive DNA data place sugarcane, *Miscanthus*, and sorghum in the subfamily Panicoideae, within the grass family (Fig. 1.2). This subfamily is one of the most distinctive groups within the grasses, having been recognized initially by Robert Brown (1810, 1814) based on its two-flowered spikelets, with the upper flower being bisexual and the lower flower staminate or sterile.

3 The Tribe Andropogoneae

3.1 Molecular Phylogenetics

Sugarcane, *Miscanthus*, and sorghum belong to the tribe Andropogoneae, a group that includes 85–90 genera (Clayton and Renvoize 1986). All molecular phylogenetic studies of Andropogoneae (Bomblies and Doebley 2005; Chen et al. 2009; Hodkinson et al. 2002a; Kellogg 2000a; Lukens and Doebley 2001; Mathews et al. 2002; Skendzic et al. 2007; Spangler et al. 1999) show that it is monophyletic and that the genus *Arundinella* (formerly the type genus of the tribe Arundinelleae) is sister to all Andropogoneae. The common ancestor of *Arundinella* plus all other Andropogoneae is estimated to have lived about 19 million years (My) ago (19.1 ± 4.5; Vicentini et al. 2008).

All studies also find good support for linking *Zea* plus *Tripsacum*, and for a clade composed of *Bothriochloa*, *Dichanthium*, and *Capillipedium*; both these groups are expected based on considerable previous work (e.g., deWet and Harlan 1974; Harlan and deWet 1963; Hitchcock 1950; Mangelsdorf and Reeves 1931). Other large groups that were identified in one or more studies include (1) the awned Andropogoneae (Bomblies and Doebley 2005; Kellogg 2000a; Lukens and Doebley 2001; Mathews et al. 2002), dated by Vicentini et al. (2008) at 11 My (11.4 ± 3.1); (2) "core Andropogoneae," including *Andropogon*, *Schizachyrium*, *Hyparrhenia*,

Cymbopogon, and *Heteropogon*, plus the clade of *Bothriochloa*, *Dichanthium*, and *Capillipedium* (Bomblies and Doebley 2005; Kellogg 2000a; Lukens and Doebley 2001; Mathews et al. 2002; Skendzic et al. 2007; Vicentini et al. 2008)(9.1 ± 2.7 My); (3) *Saccharum* plus *Miscanthus* (Hodkinson et al. 2002a; Mathews et al. 2002; Skendzic et al. 2007); (4) African sorghum (*S. bicolor*, *S. arundinaceum*, *S. halepense*, *S. propinquum*) (Chen et al. 2009; Hodkinson et al. 2002a; Skendzic et al. 2007).

Comparing the many studies is difficult because each includes a somewhat different set of taxa and a different set of DNA markers. Spangler et al. (1999) used the chloroplast gene *ndhF*, Kellogg (2000a) used *ndhF*, granule bound starch synthase 1 (GBSS1 or *waxy*) and morphology, Lukens and Doebley (2001) used *teosinte branched 1* (*tb1*), Mathews et al. (2002) combined *ndhF*, *waxy*, and phytochrome B (*phyB*), Bomblies and Doebley (2005) used *Leafy* (*lfy*), Hodkinson et al. (2002a) and Skendzic et al. (2007) used the Internal Transcribed Spacer of the nuclear ribosomal RNA genes (ITS) plus sequences of the intron and spacer of the chloroplast *trnL-F*, Chen et al. (2009) used ITS, and Vicentini et al. (2008) used *ndhF* and *phyB*. Spangler et al. (1999) included a particularly large sample of the Australian species of *Sorghum*, Bomblies and Doebley (2005) focused on *Zea* and *Tripsacum*, Hodkinson et al. (2002a) included many species of *Saccharum* and *Miscanthus*, and Chen et al. (2009) focused on *Microstegium*. In these studies, other members of Andropogoneae were included simply as placeholders. Plant material and DNA was shared among the authors of several studies, so that studies from different labs used some of the same plant accessions; for example, all sequences of *Capillipedium parviflorum* were produced from a single plant. This is useful, in that any conflict between gene trees cannot be due to misidentification or confusion of specimens. On the other hand, it means that in some cases our view of the history of an entire species is determined by the DNA sequences of a single plant.

More problematical for comparison among studies is the lack of resolution of the trees. No study—even Mathews et al. (2002), which used more base pairs of DNA than any of the others—was able to resolve the early radiation of the Andropogoneae. Few mutations were found to link any of the clades, and many genera remain unplaced relative to each other. This sort of phylogeny, with short internal branches (few mutations) and longer terminal branches, is notoriously difficult to resolve and generally requires large amounts of DNA sequence for large numbers of taxa (e.g., Baurain et al. 2007; Jian et al. 2008; Rokas and Carroll 2005; Wurdack and Davis 2009), an approach that has yet to be tried for the Andropogoneae. An attempt to synthesize the available trees for this chapter by using a supertree approach produced a tree that was almost entirely unresolved (not shown).

The poor phylogenetic resolution of Andropogoneae, and general lack of appropriate taxon sampling, means that we do not know precisely where *Saccharum*, *Miscanthus* and *Sorghum* fall within the tribe and what their closest relatives are. Based on their morphology and molecular data, they clearly belong in the awned Andropogoneae, a clade that includes about two thirds of the genera of the tribe. Molecular data also show that they fall outside the core Andropogoneae, and thus the genera of that group can be ruled out as near relatives. Some studies hint at a

clade that includes both *Saccharum* and *Sorghum* (Bomblies and Doebley 2005; Hodkinson et al. 2002a; Skendzic et al. 2007), but the group is not strongly supported. Early studies linking the two genera included too few taxa to evaluate relationships rigorously (Al-Janabi et al. 1994; Hamby and Zimmer 1988; Sobral et al. 1994). Other genera that have been linked to the *Saccharum*/*Sorghum* group by one or more studies include *Cleistachne*, *Microstegium*, *Miscanthus*, and *Sorghastrum*. Skendzic et al. (2007) provide data on several species of *Sorghastrum*, but as with all other studies, their relationship to *Sorghum* and to other Andropogoneae is ambiguous.

3.2 Morphological Evolution

Members of Andropogoneae have paired spikelets, one of which is sessile and one of which is pedicellate, although this characteristic is shared with many other Panicoideae (Kellogg 2000a; Zanotti et al. 2010). The ancestral condition for the tribe is for the lower flower of the sessile spikelet to be staminate, and the lemmas acute, lacking awns. Anatomically, epidermal papillae appear to have been ancestrally absent, and costal short cells in long rows (Watson and Dallwitz 1992). In addition, in most species the rachis (inflorescence stalk) breaks up at maturity. This character appears in the common ancestor of all members of the tribe except for *Arundinella*. The disarticulating rachis is lost independently in several genera, including *Miscanthus* and sorghum, although in sugarcane the lateral branches disarticulate as well. It also appears that hardened glumes, as exhibited by sorghum and many other genera, originated at this same point in the phylogeny, but this characteristic was lost later in evolutionary time.

All Andropogoneae use the C_4 photosynthetic pathway, and use NADP-ME as a decarboxylating enzyme. Associated with this C_4 subtype, the vascular bundles have a single sheath (Hattersley and Watson 1975). This photosynthetic pathway might constitute a synapomorphy for the Andropogoneae (Christin et al. 2008), or might have been derived earlier, possibly at the origin of the Panicoideae (Vicentini et al. 2008). The optimization of this character on the phylogeny depends heavily on the taxa included in the tree and also on the particular model of evolution used.

The earliest lineages of Andropogoneae, including *Coix* and the *Zea-Tripsacum* clade, lacked awns on the lemmas. Most phylogenies suggest that these awnless lineages form a paraphyletic grade, rather than a clade. In contrast, the awned Andropogoneae, a group that includes about two thirds of the genera (ca. 55) appear to be monophyletic, based on all molecular analyses to date (see above). (Some analyses also include *Coix* here even though it is clearly awnless.) The awn is usually twisted and hygroscopic, and is borne on the lemma. Awns are thought to be adaptations for seed dispersal (Elbaum et al. 2007; Garnier and Dajoz 2001; Peart 1979, 1981, 1984; Peart and Clifford 1987), although they also affect seed provisioning (e.g., Li et al. 2006; Motzo and Giunta 2002), and possibly also drought stress (Abebe et al. 2010). Awns are lost in some taxa, including a few relatives of sugarcane.

3.3 Chromosomal Evolution

The ancestral base chromosome number (n) for the Andropogoneae is most likely 10 (Spangler et al. 1999; Wilson et al. 1999). The subfamily Panicoideae has three clades, one with $x=9$, one with $x=10$, and the Andropogoneae; the latter two are sisters (Christin et al. 2008; Vicentini et al. 2008), further supporting 10 as the base number. The earliest diverging genus in Andropogoneae is *Arundinella*, for which chromosome numbers of $2n=16$ (Basappa and Muniyamma 1981), 18 (Christopher and Samraj 1985; Mehra 1982), 20 (Mehra 1982; Norrmann et al. 1994; Pohl and Davidse 1971; Sahni and Bir 1985), 24 (Mehra 1982; Rudyka 1990), 34 (Mehra 1982; Sinha et al. 1990), 40 (Christopher and Samraj 1985), and 60 (Mehra 1982) have been reported, although not all counts have been confirmed. Most other genera of the tribe have chromosome numbers that are multiples of 10, although a few include multiples of nine.

One popular idea is that the ancestor of the tribe had a $2n$ number of 10, and thus a haploid number of 5 (Celarier 1956; Garber 1950). This creates some appealing arithmetic to explain the origin of the two genomes of *Zea*. Proponents of this idea point to the handful of species in Andropogoneae that exhibit $n=5$ (*Coix aquatica*, and members of *Sorghum* subg. *Parasorghum* (=*Sarga*)). However, the $n=5$ species of *Sorghum* are clearly derived in the phylogeny (Spangler et al. 1999), pointing to a secondary reduction in chromosome number. This evidence is further supported by the finding (Paterson et al 2004, 2009) that the *Sorghum bicolor* ($n=10$) genome has not experienced genome duplication or paleopolyploidy in 70 My or more, ruling out the possibility of formation of its $n=10$ karyotype from $n=5$ sorghums. Price et al. (2005) noted that the direction of evolution between $n=10$ and $n=5$ sorghums is ambiguous, but their analysis did not include any other Andropogoneae. Garber (1950) and Price et al. (2005) found that chromosomes of the $x=5$ species were considerably larger than those of the $x=10$ taxa; the $x=5$ sorghums also have a higher 2C DNA content than the $x=10$ species (Price et al. 2005). These observations raise the possibility that $x=5$ represents an intriguing chromosomal fusion event, rather than the ancestral base chromosome number. One can imagine that a burst in retrotransposon activity could have led simultaneously to the abrupt increase in genome size seen in this group, as well as the genome rearrangements that led to five large rather than ten small chromosomes. Price et al. (2005) correct the published data on *S. leiocladum*, noting that this species is actually $n=5$ rather than $n=10$ as reported by Garber (1950) and cited by Spangler et al. (1999). They also note that the report of $n=5$ for *S. nitidum* is almost certainly an error, and that multiple accessions of this species are $n=10$.

Wilson et al. (1999) suggested an ancestral number of $x=8$ for maize, but this number is quite rare in Andropogoneae. One species of *Arundinella* is reported with this base number (Basappa and Muniyamma 1981), and *Chasmopodium*, with two species living in West Africa, is reported to have $n=8$; *Chasmopodium* has not been included in any phylogeny to date. Thus, if maize were convincingly shown to have arisen from an $x=8$ ancestor, it would reflect a highly unusual set of chromosomal rearrangements.

To summarize the available chromosomal data, by far the majority of Andropogoneae and their sister clade in Paniceae have a chromosome base number of 10, not 5. A handful of taxa have chromosome numbers less than $x = 10$. For convincing phylogenetic evidence that these represent the ancestral state, these taxa would have to be sisters to all other members of the tribe. *Coix* species sometimes appear near the base of the tribe, but never with strong support, and *Sorghum* subg. *Parasorghum* is always found to be derived. In addition, the genera *Sorghum* and *Arundinella* (if not others) would have to be nonmonophyletic. Because neither of these conditions holds, the phylogeny provides no evidence for $n = 5$ as being ancestral. The evidence from chromosome size and DNA content is also equivocal. The fact that the *Sorghum* species with $x = 5$ have larger chromosomes could also indicate that they are derived. The issue will only be resolved by genomic studies on the taxa with chromosome numbers other than $x = 10$.

4 Phylogeny of *Saccharum* and *Miscanthus*

Within the awned Andropogoneae is a group of species, including *Saccharum* and *Miscanthus*, characterized by having sessile and pedicellate spikelets alike in form and sex expression. Because of the lack of differentiation between the sessile and pedicellate spikelets, Clayton and Renvoize (1986) postulated that these were the most primitive of the Andropogoneae, but this hypothesis has not been supported by molecular phylogenetic data.

The "*Saccharum* complex" was defined originally by Mukherjee (1957) to include *Narenga*, *Sclerostachya*, *Erianthus* sect. *Ripidium*, and *Saccharum*. In most species in this complex, the main axis of the inflorescence is tough and does not break up at maturity; the lateral branches, however, disarticulate between the spikelet pairs. The inflorescence axis and branches are covered with long hairs, as is the base of the spikelet (callus). Although awned lemmas are common in this group, awns fail to develop in some species.

Mukherjee (1957) did not include *Miscanthus* in his original delimitation of the *Saccharum* group, but it was later added (Daniels and Williams 1975). *Miscanthus* is morphologically similar to *Saccharum*, but the sessile spikelet is actually on a short pedicel, and the lateral inflorescence branches do not break up at maturity.

The most comprehensive molecular phylogenetic study to investigate the "*Saccharum* complex" (including *Miscanthus*) is that of Hodkinson et al. (2002a). This study included multiple species of *Saccharum* and *Miscanthus*, as well as representatives of *Erianthus*, *Eulalia*, *Pogonatherum*, *Imperata*, *Narenga*, and *Spodiopogon*. As with all other phylogenetic studies in the group, the relationships are mostly weakly supported, and the results are somewhat inconclusive.

Figure 1.3 summarizes the ITS phylogeny from Hodkinson et al. (2002a) showing only groups that receive some support from their parsimony bootstrap analysis. A group corresponding to the genus *Saccharum* in the strict sense (sensu stricto, or s.s.) is well supported by the ITS sequences, and also when the *trnL-F* data are

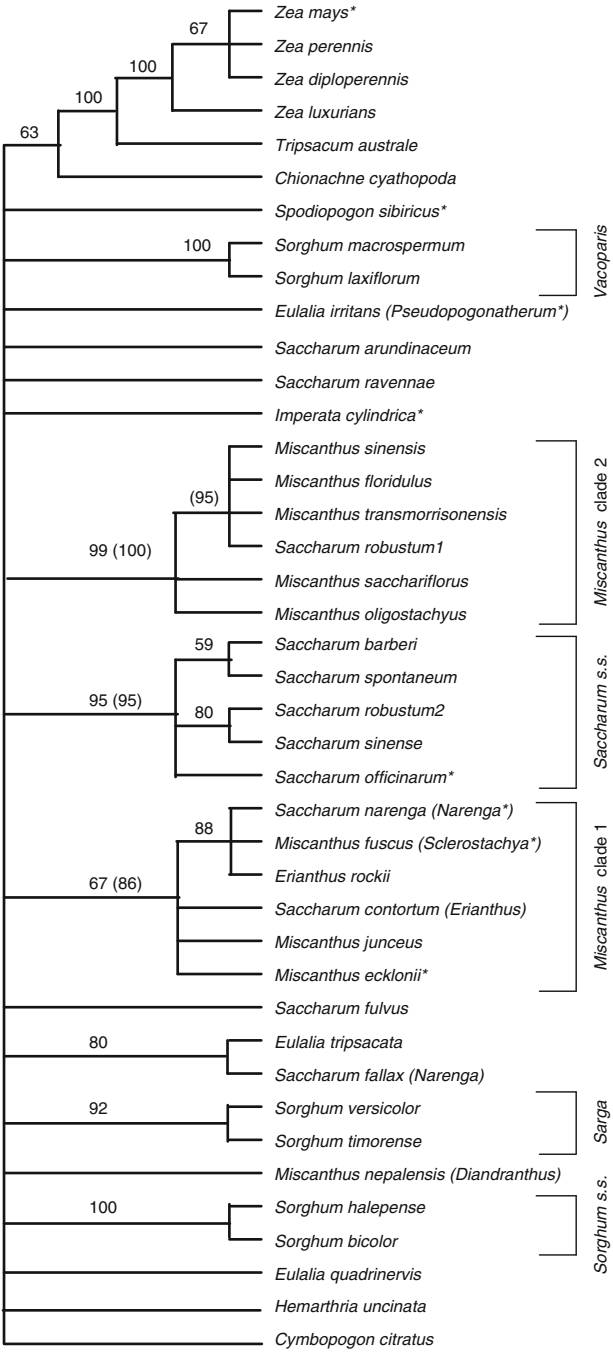

Fig. 1.3 Phylogeny of *Saccharum* and its relatives based on ITS sequences, redrawn from Hodkinson et al. (2002a). Numbers above branches are parsimony bootstrap values. Branches with less than 50 % bootstrap support are collapsed in this figure. Numbers in parentheses refer to support values obtained when the ITS data were combined with data for *trnL-F*. *Sorghum robustum* 1 and 2, and *Miscanthus floridulus* 1 and 2 are distinct paralogues from the same plant. Type species are marked with an asterisk

added (number in parentheses). There are two distinct clades of *Miscanthus* species. *Miscanthus* clade 1 includes the type species of *Miscanthus*, *M. ecklonii*, as well as species previously assigned to *Erianthus*, *Sclerostachya*, and *Narenga*; this group is only moderately supported by ITS data, but receives stronger support when the *trnL-F* data were added. *Miscanthus* clade 2 includes several species of *Miscanthus* plus one ITS paralogue from *Saccharum robustum*. *Miscanthus* × *giganteus* is a triploid derivative of *M. sinensis* and *M. sacchariflorus* (Hodkinson et al. 2002b), so could also be assigned to this clade.

Several earlier studies produced preliminary results that are consistent with those of Hodkinson et al. (2002a). Nair et al. (1999) used RAPD markers and found a group corresponding to *Saccharum* s.s., and another similar to *Miscanthus* clade 1; the analysis was phenetic, however (using Unweighted Pair Group Method with Arithmetic Means, or UPGMA), and cannot be directly compared to phylogenetic studies. Besse et al. (1997) used RFLP data to show that seven species of *Erianthus* were distinct from two species of *Saccharum*, and Selvi et al. (2006) likewise found a clear distinction between *Saccharum* species and *Erianthus* using AFLPs. Bacci et al. (2001) generated an ITS phylogeny of sugarcane and its relatives, and also found a clade corresponding to *Saccharum* s.s.

Within *Saccharum*, six species are commonly recognized: *S. officinarum*, *S. robustum*, *S. spontaneum*, *S. sinense*, *S. barberi*, and *S. edule*, although the grass species index at the Royal Botanic Gardens, Kew lists 37 names (http://www.kew.org/data/grassbase/index.html). Relationships among the species are not well resolved by available data. A study of DNA sequences from 18 chloroplast regions suggested that *S. spontaneum* is sister to the remaining species (Takahashi et al. 2005). This contradicts the results of several other studies, including the relationships shown in Fig. 1.3. However, an extensive review of Saccharum literature (Chap. 3 of this volume) supports this relationship, also generally viewing *S. sinense*, *S. barberi*, and *S. edule* as forms of *S. officinarum* modified by interspecific hybridization with *S. spontaneum*. Few of the studies have used accepted phylogenetic methods, and none has attempted to dissect the complex reticulate history of the *Saccharum* species using multiple single copy nuclear genes. It is virtually certain that the relationships within the genus *Saccharum* are not strictly divergent, and hence attempts to represent them as a tree are probably misleading.

5 Phylogeny of Sorghum

Sorghum has a highly branched panicle that is superficially quite different from the set of long racemes of *Saccharum*. The glumes are hardened, as is characteristic of many Andropogoneae. The sessile and pedicellate spikelets are quite different morphologically, with the pedicellate spikelet generally much smaller than the sessile one and either staminate or sterile.

Sorghum has been the subject of several molecular phylogenies, of which two have included all or nearly all the species (Dillon et al. 2007; Ng'uni et al. 2010).

The genus has also been the subject of a recent monograph (Spangler 2003). Thus, we know what species are in the genus and what their relationships are. Ng'uni et al. (2010) used noncoding regions of the chloroplast and the ITS; Dillon et al. (2007) added data on a nuclear gene (*Adh1*) to her previous data on ITS1 and *ndhF* (Dillon et al. 2004).

Both Dillon et al. (2007) and Ng'uni et al. (2010) found two well-supported major clades within *Sorghum*. Clade 1 includes *Sorghum bicolor* and its close relatives *S. halepense*, *S. propinquum*, *S. arundinaceum*, *S. almum*, and *S. drummondii*, supporting earlier work (Sun et al. 1994). (Note that DeWet (1978) places most species of this group except *S. halepense* and *S. propinquum* in the synonymy of *S. bicolor*; however, he does not deal with *S. almum*.) These species are all African except for *S. propinquum*, which is Asian, and are all part of the secondary gene pool of grain sorghum (*S. bicolor*) (Price et al. 2006). Also in Clade 1 are *Sorghum macrospermum* and *S. laxiflorum*, two Australian species that are clearly sisters and are more closely related to cultivated sorghum than any of the other Australian species (Dillon et al. 2004).

Clade 2 within *Sorghum* includes the 17 species have been assigned to subgenera *Stiposorghum* and *Parasorghum*, all Australian species that can be recognized easily by their bearded nodes (Snowden 1935). The Australian species of *Sorghum* are not interfertile with cultivated sorghum, and pollen from members of one group will not germinate on the stigmas of the other (Garber 1950; Hodnett et al. 2005; Price et al. 2006). All phylogenetic studies support the Australian clade, but show that species of the two subgenera are intermixed (Dillon et al. 2007; Ng'uni et al. 2010; Spangler et al. 1999; Sun et al. 1994). Spangler (2003) evaluated morphological similarities among the 17 species and concluded that there were only seven, although Dillon et al. (2007) argued for reinstating several (Table 1.1).

The trees differ in the placement of *S. nitidum*, a widespread species of Australia and Asia. Ng'uni et al. (2010) place it sister to Clade 2, whereas Dillon et al. (2007) place it within that clade, sister to *S. leiocladum*. Sun et al. (1994) place *S. nitidum* as sister to the African sorghums, and Spangler et al. (1999) place it sister to *S. laxiflorum*. Resolution of its placement will probably require sampling of multiple individuals from different parts of its range.

Cleistachne sorghoides has been included in some studies but not others. The plants look similar to cultivated sorghum (hence the specific epithet) but lack the pedicellate spikelet; a close association of the two genera has been postulated since the 19th century (Hackel 1889). Both Sun et al. (1994) and Dillon et al. (2004) placed *Cleistachne* among the Australian sorghums, a somewhat surprising result because *Cleistachne* is African. However, Dillon et al. (2007) place *Cleistachne* sister to the clade made up of subg. *Sorghum* and *Heterosorghum*. The difference likely reflects the signal from *Adh1*, which was only included by Dillon et al. (2007).

Monophyly of sorghum in its traditional sense (i.e., clades 1 and 2 together) is neither proven nor disproven by available data. Dillon et al. (2007) strongly assert that *Sorghum* is monophyletic. However, both they and Ng'uni et al. (2010) begin with an assumption of monophyly, and do not include *Microstegium*, *Saccharum*, *Miscanthus*, *Erianthus*, *Imperata*, *Sorghastrum*, or any of the other taxa that could

Table 1.1 Alternate classifications of the genus *Sorghum*

Single genus classification (Dillon et al. 2007)	Three-genus classification (Spangler 2003)
Sorghum subg. *Parasorghum*	*Sarga*
Sorghum angustum	*Sarga angustum*
Sorghum ecarinatum	[not named in *Sarga*]
Sorghum interjectum	[not named in *Sarga*]
Sorghum intrans	*Sarga intrans*
Sorghum leiocladum	*Sarga leiocladum*
Sorghum nitidum	[not named in *Sarga*]
Sorghum plumosum (=*S. grande*)	*Sarga plumosum*
Sorghum purpureo-sericeum (=*S. pappii, S. deccanense, S. dimiditum*)	*Sarga purpureo-sericeum*
Sorghum timorense (=*S. amplum, S. australiense, S. brachypodum, S. brevicallosum, S. bulbosum, S. matarankense, S. mjoebergii, S. stipoideum*)	*Sarga timorense*
Sorghum trichocladum	*Sarga trichocladum*
Sorghum versicolor	*Sarga versicolor*
Sorghum subg. *Sorghum*	*Sorghum*
Sorghum bicolor	*Sorghum bicolor*
Sorghum halepense	(=*S. propinquum, S.* × *almum, S.* × *drummondii)*
Sorghum propinquum	*Sorghum halepense*
Sorghum × *almum*	*Sorghum nitidum*
Sorghum × *drummondii*	
Sorghum subg. *Chaetosorghum*	*Vacoparis*
Sorghum laxiflorum	*Vacoparis laxiflorum*
Sorghum macrospermum	*Vacoparis macrospermum*

be related to one or the other of the sorghum clades. Spangler et al (1999) and Hodkinson et al. (2002a) included a broad sample of *Sorghum* species and found that the well-supported clades might not actually particularly closely related (Fig. 1.3). However, because of the lack of resolution of the backbone of the tree, monophyly of the genus cannot be ruled out.

6 Classification of Saccharinae and Sorghinae

In an ideal world, the classification of a plant would be based directly on its position in a phylogenetic tree, such that the name provides unambiguous information on genealogical relationships. Much taxonomic effort in recent years has gone into generating such trees, and adjusting classifications to reflect current knowledge of relationships. In the case of Saccharinae and Sorghinae, unfortunately, the molecular data are inconclusive, so that current classifications rely on a mix of morphological observations (the traditional classification) and fragmentary knowledge of relationships.

Saccharinae and Sorghinae are subtribes within the larger tribe Andropogoneae. The International Code of Botanical Nomenclature (ICBN) (McNeill et al. 2007) requires only the ranks of family, genus and species, but provides intermediate ranks to be used as needed. In the grass family it is conventional to divide the family into subfamilies, the subfamilies into tribes and the tribes into subtribes. There are 12 subfamilies currently recognized (Grass Phylogeny Working Group 2001), of which the subfamily Panicoideae is one. Panicoideae is in turn divided into several tribes, one of which is Andropogoneae. The rank of subtribe is not used nearly as often as the others, and thus is much less familiar. Perhaps because of this, membership of subtribes tends to receive little attention, and changes to that membership likewise are often poorly known.

Because the application of the name Saccharinae is so variable, I need to provide a brief overview of how naming proceeds and what rules govern it. Grouping, ranking, and naming are separate activities in taxonomy (Judd et al. 2007). Grouping is the process of deciding which organisms belong together (e.g., which plants are closely enough related to be designated a species, which species belong together to form a genus, etc.). Current taxonomic practice groups organisms together if they constitute all the descendants of a single common ancestor, i.e., if together they form a clade (a monophyletic group). Grouping thus depends absolutely on a phylogenetic tree to determine which sets of organisms are monophyletic. Historically, however, grouping often depended solely on the way the organisms looked (their morphology) so that groups were made up of similar-looking organisms. These might be similar because of shared ancestry—i.e., similar morphology might indicate that they were all a member of a monophyletic group—or because of convergence—i.e., they might be grouped together even though they had little in common.

Ranking is second. Once a group is defined as monophyletic, the rank (genus, subtribe, tribe, subfamily, etc.) is determined. Unlike grouping, in which clear criteria can be applied, ranking is arbitrary and is determined by convention. Thus, a tribe, subtribe, or genus does not need to have a particular size or age or amount of morphological disparity; the ranks are simply relative. The primary criterion for ranking is thus simply nomenclatural stability. For example, the Andropogoneae could be considered a subfamily (rather than a tribe), but there is no particular reason to do so, and it remains a tribe by convention.

The third step is naming. Once a group has been identified and the rank is decided, then the taxonomist considers any name at that rank that has ever been applied to any member of that group. If the group contains more than one name, then the oldest name will be applied.

This process may be illustrated by reference to Fig. 1.3. The clade labeled *Saccharum* s.s., is a monophyletic group and hence fulfills the grouping criterion; the rank stays at the level of genus for stability. Included in the group are the species *S. barberi*, *S. edule*, *S. robustum*, *S. spontaneum*, *S. sinense*, and *S. officinarum*, which were named in 1925, 1842, 1946, 1771, 1818, and 1753, respectively. The oldest name is *Saccharum officinarum*, applied by Linnaeus to the cultivated sugarcane. Therefore, *S. officinarum* is the type species for *Saccharum*, and any genus-level group that includes *S. officinarum* must be called *Saccharum*.

Likewise, the type species of *Miscanthus* is *M. ecklonii*, which is in *Miscanthus* clade 1. Also in *Miscanthus* clade 1 are the type species of *Narenga* (*S. narenga*) and of *Sclerostachya* (*S. fuscum*, which is another name for *Miscanthus fuscus*). Because the dates for *Miscanthus*, *Narenga*, and *Sclerostachya* are 1856, 1889, and 1940, respectively, the names *Narenga* and *Sclerostachya* can be discarded in favor of the oldest name, *Miscanthus*.

Erianthus, however, complicates the issue. The name *Erianthus* was created in 1803 (Michaux 1803), to accommodate species that looked similar to *Saccharum*, but had awns. The type species is *E. saccharoides*, a species that is also known as *Saccharum giganteum*. If a future study placed *E. saccharoides* with the two other species of *Erianthus*, in *Miscanthus* clade 1, then the name *Erianthus* would take precedence over *Miscanthus*.

Meanwhile, "*Miscanthus*" clade 2 includes no type species and hence there are no names associated with this clade. Hodkinson et al. (2002a) place *Miscanthus* clade 2 sister to *Saccharum*, although with very little support. If future data support a strong relationship between *Miscanthus* clade 2 and *Saccharum*, then the name *Saccharum* would take precedence for the whole group and such taxa as *M. sinensis* would be renamed as species of *Saccharum*.

To summarize what we know about generic limits, sugarcane and its immediate relatives in the genus *Saccharum* form a clade. The names *Narenga* and *Sclerostachya* clearly do not refer to distinct evolutionary lineages and should be dropped. The limits of the genera *Miscanthus* and *Erianthus* are unclear, as is their relationship to *Saccharum*.

Within *Sorghum* s.l., the taxonomic problem is only partially a grouping problem; it is also a ranking problem. All data point to the same three clades within the genus (Table 1.1). If the genus *Sorghum* is considered in its broad sense, then the three clades remain subgenera (or sections; the choice is arbitrary and names have been validly published at both ranks). Subgenus *Eu-sorghum* should be subg. *Sorghum* because the ICBN requires dropping the prefix "eu" for the type subgenus (McNeill et al. 2007). *Chaetosorghum* and *Heterosorghum* can be merged as recommended by Dillon et al. (2007). The two were created simultaneously by Garber (1950), so the choice of which name to use is arbitrary; I suggest *Chaetosorghum* here.

Parasorghum and *Stiposorghum* should also be merged; the name *Parasorghum* is older than *Stiposorghum* (1935 vs. 1950), so takes precedence. However, if the genus is split up, the same three clades become *Sorghum* s.s, *Sarga*, and *Vacoparis* (Spangler 2003) (Table 1.1). At issue is a grouping question at the next higher level of the taxonomic hierarchy; it is not clear whether *Sorghum* s.l. is monophyletic. Note also that Spangler (2003) had a broader concept of *S. bicolor* than other authors (Table 1.1).

The phylogenetic and taxonomic problems just described for the genera are echoed in the difficulties with the subtribes. The genera of Saccharinae and Sorghinae, as delimited by Clayton and Renvoize (1986), are listed in Table 1.2. Because of the apparent close relationship between members of the *Saccharum* complex and *Sorghum*, it is possible that the two groups should be merged. Addressing first the issue of grouping, several genera currently included in the

Table 1.2 Classification of Saccharinae and Sorghinae

Clayton and Renvoize (1986)	This paper, classification I	This paper, classification II
Subtribe Saccharinae	Subtribe Saccharinae	Subtribe Saccharinae
Eriochrysis	***Erianthus***	*Asthenochloa*
Eulalia	*Eriochrysis*	***Cleistachne***
Eulaliopsis	*Eulaliopsis*	***Erianthus***
Homozeugos	*Homozeugos*	*Eriochrysis*
Imperata	***Imperata***	*Euclasta*
Lophopogon	***Leptatherum***	*Eulaliopsis*
Microstegium	*Lophopogon*	*Hemisorghum*
Miscanthus	***Miscanthus***	*Homozeugos*
Pogonatherum	***Pogonatherum***	***Imperata***
Polliniopsis	*Polytrias*	***Leptatherum***
Polytrias	***Saccharum***	*Lophopogon*
Saccharum		***Miscanthus***
Spodiopogon		***Pogonatherum***
		Polytrias
Subtribe Sorghinae	Subtribe Sorghinae	*Pseudosorghum*
Asthenochloa	*Asthenochloa*	***Saccharum***
Bothriochloa	***Cleistachne***	***Sorghastrum***
Capillipedium	*Euclasta*	***Sorghum* s.l.**
Chrysopogon	*Hemisorghum*	*Spathia*
Cleistachne	*Pseudosorghum*	
Dichanthium	***Sorghastrum***	
Euclasta	***Sorghum* s.l.**	
Hemisorghum	*Spathia*	
Pseudichanthium		
Pseudosorghum		
Sorghastrum		
Sorghum		
Spathia		
Vetiveria		

Molecular data are available for taxa shown in boldface

Saccharinae and Sorghinae can be excluded based on molecular data. *Bothriochloa*, *Capillipedium*, and *Dichanthium* clearly are more closely related to the core Andropogoneae. Likewise, data presented by Hodkinson et al. (2002a) suggest that *Spodiopogon* may be more closely related to the core Andropogoneae and should perhaps be excluded. The genus *Microstegium* is not monophyletic (Chen et al. 2009; Spangler et al. 1999). While *M. nudum* is related to *Saccharum* and *Sorghum*, *M. vimineum* is not. *M. nudum* has been assigned to the genus *Leptatherum* (Chen et al. 2009), and *Microstegium* s.s. excluded from Saccharinae. *Polliniopsis* is considered to be part of *Leptatherum* (Chen et al. 2009).

If Saccharinae and Sorghinae are merged, then the name Saccharinae will take precedence. Saccharinae was first described in 1844. While Clayton and Renvoize (1986) indicate that Sorghinae was described in 1836, a check of the original publication shows that the name was not validly published (Soreng et al. 2008); the name

was formally, if inadvertently, validated by Clayton and Renvoize (1986) so that 1986 becomes the effective publication date. Since it is later than 1844, Saccharinae has priority.

Sufficient data are not available for a definitive answer to what should be included in Saccharinae. However, there are two possibilities that could be suggested here. Possibility 1 would keep the two tribes separate until additional phylogenetic data can be assembled to test whether they truly form a clade. Possibility 2 would be to subsume Sorghinae into Saccharinae, on the assumption that they will eventually be shown to form a clade, but that neither is monophyletic on its own. Saccharinae would then include Saccharinae s.s. plus parts of the former Sorghinae.

7 Summary and Conclusions

In summary, the phylogenetic data support several conclusions. *Saccharum* s.s. and *Sorghum* s.s. are each monophyletic. *Narenga* and *Sclerostachya* are not distinct genera. Within *Sorghum*, subgenus *Stiposorghum* can be merged into *Parasorghum*, and *Heterosorghum* into *Chaetosorghum*. Sugarcane, *Miscanthus*, and sorghum are members of the awned Andropogoneae, and as such are relatively distantly related to the *Zea*/*Tripsacum* clade. They are also not closely related to the core Andropogoneae. Neither the Saccharinae nor the Sorghinae as defined by Clayton and Renvoize (1986) is monophyletic.

However, despite multiple phylogenetic studies, many problems remain in the phylogeny and classification of the Saccharinae and Sorghinae. The circumscription of the subtribe(s) is unclear; the data hint at the very real possibility that Saccharinae should in fact be expanded to include Sorghinae. The clade most closely related to *Saccharum* is uncertain, and it is not clear what species should be included in *Miscanthus* and *Erianthus*, or whether either name should be retained at all. *Sorghum* s.l. may or may not be monophyletic.

As with many such phylogenetic problems, the solution will require collecting much more sequence data for many more species. It will be important to include all species and genera that might possibly be involved in the phylogenetic problem, and to expand the study beyond any single genus.

Acknowledgment I thank Editor Andrew Paterson for helpful comments on the manuscript.

References

Abebe T, Melmaiee K, Berg V, Wise RP (2010) Drought response in the spikes of barley: gene expression in the lemma, palea, awn, and seed. Func Integr Genomics 10:191–205

Al-Janabi SM, McClelland M, Petersen C, Sobral BWS (1994) Phylogenetic analysis of organellar DNA sequences in the Andropogoneae: Saccharinae. Theor Appl Genet 88:933–944

Bacci M Jr, Miranda VFO, Martins VG, Figueira AVO, Lemos MV, Pereira JO, Marino CL (2001) A search for markers of sugarcane evolution. Genet Mol Biol 24:169–174

Basappa GP, Muniyamma M (1981) Reproduction in two species of *Arundinella* Raddi, Poaceae. Proc Indian Acad Sci Sect B, Biol Sci 90:477–483

Baurain D, Brinkmann H, Philippe H (2007) Lack of resolution in the animal phylogeny: closely spaced cladogeneses or undetected systematic errors? Mol Biol Evol 24:6–9

Besse P, McIntyre CL, Berding N (1997) Characterisation of *Erianthus* sect. *Ripidium* and *Saccharum* germplasm (Andropogoneae-Saccharinae) using RFLP markers. Euphytica 93:283–292

Bomblies K, Doebley JF (2005) Molecular evolution of FLORICAULA/LEAFY orthologs in the Andropogoneae (Poaceae). Mol Biol Evol 22:1082–1094

Brown R (1810) Prodromus florae Novae Hollandiae. J. Johnson & Co., London

Brown R (1814) A voyage to Terra Australis. G. & W. Nicol, London

Campbell CS, Kellogg EA (1987) Sister group relationships of the Poaceae. In: Soderstrom TR, Hilu KW, Campbell CS, Barkworth ME (eds) Grass systematics and evolution. Smithsonian Institution, Washington, DC, pp 217–224

Celarier RP (1956) Additional evidence for five as the basic chromosome number of the Andropogoneae. Rhodora 58:135–143

Chen C-H, Veldkamp JF, Kuoh C-S, Tsai C-C, Chiang Y-C (2009) Segregation of *Leptatherum* from *Microstegium* (Andropogoneae, Poaceae) confirmed by internal transcribed spacer DNA sequences. Blumea 54:175–180

Christin PA, Besnard G, Samaritani E, Duvall MR, Hodkinson TR, Savolainen V, Salamin N (2008) Oligocene CO_2 decline promoted C_4 photosynthesis in grasses. Curr Biol 18:37–43

Christopher J, Samraj P (1985) Chromosome number reports LXXXVI. Taxon 34:159–164

Clayton WD, Renvoize SA (1986) Genera graminum: grasses of the world. Her Majesty's Stationery Office, London

Daniels J, Williams CA (1975) The origin of the genus *Saccharum*. ISSCT Sugarcane Breed Newsl 36:24–39

deWet JMJ (1978) Systematics and evolution of *Sorghum* sect. *Sorghum* (Gramineae). Am J Bot 65:477–484

deWet JMJ, Harlan JR (1974) Tripsacum-maize interaction: a novel cytogenetic system. Genetics 78:493–502

Dillon SL, Lawrence PK, Henry RJ, Price HJ (2007) *Sorghum* resolved as a distinct genus based on combined ITS1, *ndh*F and *Adh*1 analyses. Plant Syst Evol 268:29–43

Dillon SL, Lawrence PK, Henry RJ, Ross L, Price HJ, Johnston JS (2004) *Sorghum laxiflorum* and *S. macrospermum*, the Australian native species most closely related to the cultivated *S. bicolor* based on ITS1 and *ndhF* sequence analysis of 25 *Sorghum* species. Plant Syst Evol 249:233–246

Elbaum R, Zaltzman L, Burgert I, Fratzl P (2007) The role of wheat awns in the seed dispersal unit. Science 316:884–886

Garber ED (1950) Cytotaxonomic studies in the genus *Sorghum*. Univ Calif Publ Bot 23:283–362

Garnier LKM, Dajoz I (2001) Evolutionary significance of awn length variation in a clonal grass of fire-prone savannas. Ecology 82:1720–1733

Grass Phylogeny Working Group (2001) Phylogeny and subfamilial classification of the Poaceae. Ann Missouri Bot Gard 88:373–457

Hackel E (1889) Andropogoneae. In: DeCandolle A, DeCandolle C (eds) Monographiae phanerogamarum. G. Masson, Paris

Hamby RK, Zimmer EA (1988) Ribosomal RNA sequences for inferring phylogeny within the grass family (Poaceae). Plant Syst Evol 160:29–37

Harlan JR, deWet JMJ (1963) The compilospecies concept. Evolution 17:497–501

Hattersley PW, Watson L (1975) Anatomical parameters for predicting photosynthetic pathways of grass leaves: The "maximum lateral cell count" and the "maximum cells distant count". Phytomorphology 25:325–333

Hitchcock AS (1950) Manual of the grasses of the United States. U.S. Government Printing Office, Washington, DC

Hodkinson TR, Chase MW, Lledó MD, Salamin N, Renvoize SA (2002a) Phylogenetics of *Miscanthus*, *Saccharum* and related genera (Saccharinae, Andropogoneae, Poaceae) based on DNA sequences from ITS nuclear ribosomal DNA and plastid *trnL* intron and *trnL-F* intergenic spacers. J Plant Res 115:381–392

Hodkinson TR, Chase MW, Takahashi C, Leitch IJ, Bennett MD, Renvoize SA (2002b) The use of DNA sequencing (ITS and *trnL-F*), AFLP, and fluorescent in situ hybridizaiton to study allopolyploid *Miscanthus* (Poaceae). Am J Bot 89:279–286

Hodnett GL, Burson BL, Rooney WL, Dillon SL, Price HJ (2005) Pollen-pistil interactions result in reproductive isolation between *Sorghum bicolor* and divergent *Sorghum* species. Crop Sci 45:1403–1409

Jian S, Soltis PS, Gitzendanner MA, Moore MJ, Li R, Hendry TA, Qiu Y-L, Dhingra A, Bell CD, Soltis DE (2008) Resolving an ancient, rapid radiation in Saxifragales. Syst Biol 57:38–57

Judd WS, Campbell CS, Kellogg EA, Stevens PF, Donoghue MJ (2007) Plant systematics: a phylogenetic approach. 3rd edition., 3rd edn. Sinauer Asociates, Sunderland, Massachusetts

Kellogg EA (2000a) Molecular and morphological evolution in Andropogoneae. In: Jacobs SWL, Everett JE (eds) Grasses: systematics and evolution. CSIRO, Melbourne, pp 149–158

Kellogg EA (2000b) The grasses: a case study in macroevolution. Annu Rev Ecol Syst 31:217–238

Kellogg EA, Linder HP (1995) Phylogeny of Poales. In: Rudall PJ, Cribb PJ, Cutler DF, Humphries CJ (eds) Monocotyledons: systematics and evolution. Royal Botanic Gardens, Kew, pp 511–542

Li X, Wang H, Li H, Zhang L, Teng N, Lin Q, Wang J, Kuang T, Li Z, Li B, Zhang A, Lin J (2006) Awns play a dominant role in carbohydrate production during the grain-filling stages in wheat (*Triticum aestivum*). Physiol Plant 127:701–709

Lukens L, Doebley J (2001) Molecular evolution of the *teosinte branched* gene among maize and related grasses. Mol Biol Evol 18:627–638

Mangelsdorf PC, Reeves RG (1931) Hybridization of maize, *Tripsacum* and *Euchlaena*. J Hered 22:329–343

Mathews S, Spangler RE, Mason-Gamer RJ, Kellogg EA (2002) Phylogeny of Andropogoneae inferred from phytochrome B, GBSSI, and *ndhF*. Int J Plant Sci 163:441–450

McNeill J, Barrie FR, Burdet HM, Demoulin V, Hawksworth DL, Marhold K, Nicolson DH, Prado J, Silva PC, Skog JE, Wiersema JH, Turland NJ (eds) (2007) International code of botanical nomenclature (Vienna Code). Ruggell, Gantner

Mehra PN (1982) Cytology of East Indian grasses. Chandigarh, India

Michaux A (1803) Flora Boreali-Americana. Levrault, Paris

Motzo R, Giunta F (2002) Awnedness affects grain yield and kernel weight in near-isogenic lines of durum wheat. Aust J Agric Res 53:1285–1293

Mukherjee SK (1957) Origin and distribution of *Saccharum*. Bot Gaz 119:55–61

Nair NV, Nair S, Sreenivasan TV, Mohan M (1999) Analysis of genetic diversity and phylogeny in Saccharum and related genera using RAPD markers. Genet Res Crop Evol 46:73–79

Ng'uni D, Geleta M, Gatih M, Bryngelsson T (2010) Phylogenetic analysis of the genus *Sorghum* based on combined sequence data from cpDNA regions and ITS generate well-supported trees with two major lineages. Ann Bot 105:471–480

Norrmann GA, Quarín CL, Killeen TJ (1994) Chromosome numbers in Bolivian grasses (Gramineae). Ann Missouri Bot Gard 81:768–774

Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc Natl Acad Sci U S A 101:9903–9908

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannag M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob-ur-Rahman WD, Westhoff P, Mayer KFX, Messing J, Rokhsar DS (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556

Peart MH (1979) Experiments on the biological significance of the morphology of seed-dispersal units in grasses. J Ecol 67:843–863
Peart MH (1981) Further experiments on the biological significance of the morphology of seed-dispersal units in grasses. J Ecol 69:425–436
Peart MH (1984) The effects of morphology, orientation and position of grass diaspores on seedling survival. J Ecol 72:437–453
Peart MH, Clifford HT (1987) The influrence of diaspore morphology and soil-surface properties on the distribution of grasses. J Ecol 75:569–576
Pohl RW, Davidse G (1971) Chromosome numbers of Costa Rican grasses. Brittonia 23:293–324
Price HJ, Dillon SL, Hodnett G, Rooney WL, Ross L, Johnston JS (2005) Genome evolution in the genus *Sorghum* (Poaceae). Ann Bot 95:219–227
Price HJ, Hodnett GL, Burson BL, Dillon SL, Stelly DM, Rooney WL (2006) Genotype dependent interspecific hybridization of *Sorghum bicolor*. Crop Sci 46:2617–2622
Rokas A, Carroll SB (2005) More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. Mol Biol Evol 22:1337–1344
Rudyka EG (1990) Chromosome numbers of vascular plants from the various regions of the USSR. Bot Zhurn 75:1783–1786
Sahni M, Bir SS (1985) SOCGI plants chromosome number reports - III. J Cytol Genet 20:205–206
Selvi A, Nair NV, Noyer JL, Singh NK, Balasundaram N, Bansal KC, Koundal KR, Mohapatra T (2006) AFLP analysis of the phenetic organization and genetic diversity in the sugarcane complex, *Saccharum* and *Erianthus*. Genet Res Crop Evol 53:831–842
Sinha RRP, Bhardwaj AK, Singh RK (1990) SOCGI plant chromosome number reports - IX. J Cytol Genet 25:140–143
Skendzic EM, Columbus JT, Cerros-Tlatilpa R (2007) Phylogenetics of Andropogoneae (Poaceae: Panicoideae) based on nuclear ribosomal internal transcribed spacer and chloroplast *trn*L-F sequences. Aliso 23:530–544
Snowden JD (1935) A classification of the cultivated Sorghums. Bull Misc Inform Royal Bot Gard Kew 5:221–255
Sobral BWS, Braga DPV, LaHood ES, Keim P (1994) Phylogenetic analysis of chloroplast restriction enzyme site mutations in the Saccharinae Griseb. subtribe of the Andropogoneae Dumort. tribe. Theor Appl Genet 87:843–853
Soreng RJ, Davidse G, Peterson PM, Zuloaga FO, Judziewicz EJ, Filgueiras TS, Morrone O (2008) Catalogue of New World Grasses (Poaceae). http://mobot.mobot.org/W3T/Search/nwgc.html#Status First published 2000 and updated frequently
Spangler R, Zaitchik B, Russo E, Kellogg E (1999) Andropogoneae evolution and generic limits in *Sorghum* (Poaceae) using *ndhF* sequences. Syst Bot 24:267–281
Spangler RE (2003) Taxonomy of *Sarga*, *Sorghum* and *Vacoparis* (Poaceae: Andropogoneae). Aust Syst Bot 16:279–299
Stevens PF (2008) Angiosperm phylogeny website. Version 9, June 2008 [and more or less continuously updated since]. 2001 onwards
Sun Y, Skinner DZ, Liang GH, Hulbert SH (1994) Phylogenetic analysis of *Sorghum* and related taxa using internal transcribed spacers of nuclear ribosomal DNA. Theor Appl Genet 89:26–32
Takahashi S, Furukawa T, Asano T, Terajima Y, Shimada H, Sugimoto A, Kadowski K (2005) Very close relationship of the chloroplast genomes among *Saccharum* species. Theor Appl Genet 110:1523–1529
Vicentini A, Barber JC, Giussani LM, Aliscioni SS, Kellogg EA (2008) Multiple coincident origins of C_4 photosynthesis in the Mid- to Late Miocene. Global Change Biol 14:2963–2977
Watson L, Dallwitz MJ (1992) Grass genera of the world. CAB International, Wallingford, CT
Wilson WA, Harrington SE, Woodman WL, Lee M, Sorrells ME, McCouch SR (1999) Inferences on the genome structure of progenitor maize through comparative analysis of rice, maize and the domesticated panicoids. Genetics 153:453–473
Wurdack KJ, Davis CC (2009) Malpighiales phylogenetics: gaining ground on one of the most recalcitrant clades in the angiosperm tree of life. Am J Bot 96:1551–1570
Zanotti CA, Pozner R, Morrone O (2010) Understanding spikelet orientation in Paniceae (Poaceae). Am J Bot 97:717–729

Chapter 2
The Gene Pool of *Sorghum bicolor* and Its Improvement

Clarissa T. Kimber, Jeff A. Dahlberg, and Stephen Kresovich

Abstract Sorghum, a genus having evolved across a wide range of environments in Africa, exhibits a great range of phenotypic diversity and numerous resistances to abiotic and biotic stresses. Sorghum is recognized as a highly productive, drought tolerant, C_4 cereal that provides humankind with food, feed, fuel, fiber, and energy, particularly in the semiarid tropics of the world.

Sorghum has been collected and conserved over the past 50 years and numerous international and national collections exist. The major collections have in excess of 40,000 accessions and much of the native diversity of cultivated sorghum is represented. However, much of the diversity of the wild races of sorghum is underrepresented in these same collections. Over the past decade, the major collections have benefited by efforts to better characterize these accessions but these efforts have not significantly increased use of the materials. Therefore, despite a significant number of collections and holdings, much of the diversity of sorghum remains untapped.

Over the past decade, tremendous progress has been made to build the molecular and genomic foundation required to increase our understanding of sorghum diversity in the genome and gene pool and, ultimately, to link this information to crop improvement. Sorghum represents the first crop genome of African origin to be sequenced (Paterson et al. Nature 457:551–556, 2009) and, through coordinated national and international efforts, high-density genetic and physical maps, extensive

C.T. Kimber (✉)
Department of Geography, Texas A&M University, 705 O&M Building, College Station, TX 77843, USA
e-mail: ctk5617@geog.tamu.edu

J.A. Dahlberg
U.S. Sorghum Checkoff Program, Lubbock, TX, USA
e-mail: jeff@sorghumcheckoff.com

S. Kresovich
Department of Biology, University of South Carolina, Columbia, SC, USA
e-mail: sk@sc.edu

A.H. Paterson (ed.), *Genomics of the Saccharinae*, Plant Genetics and Genomics: Crops and Models 11, DOI 10.1007/978-1-4419-5947-8_2,

sets of RFLP and SSR markers, association (Casa et al. Crop Sci 48:30–40, 2008) and diversity panels (Deu et al.Genome 49:168–180, 2006), nested association mapping populations, and other resources are readily available for use for scientific investigations and breeding efforts. The generation and use of these genomic resources have added to our insights about sorghum domestication and diversity. Future studies will enrich our understanding and provide increasing resolution to quantify and use both wild and domesticated sources of diversity in crop improvement.

Keywords Bottleneck • Collections • Conservation • Diversity • Domestication • Evolution • Genome • Races • Selection

1 Properties of the Species

1.1 Introduction

Having been domesticated for a variety of useful products and cultivated in a broad range of environments, sorghum exhibits a great range of phenotypic diversity. Around the world, sorghum is grown for the production of dense grain panicles (for food, feed, and/or energy), tall, thick sweet stalks (for food, feed, and/or energy), and various forage types (for feed and fuel). Through advanced genetic manipulation, striking divergence among the different forms of sorghum can occur (see Fig. 2.1). In regions where mechanical harvesting predominates, grain types tend to be short in stature (0.5–1.0 m) with a large erect stem supporting a semicompact or compact panicle. In regions where manual harvesting is still practiced and multiple plant parts are desired, the plant may be tall (3–5 m) with an open panicle.

As a member of the grass family, sorghum represents a robust, cane-like species. It has the ability to tiller, regrow following harvest, and produce a fibrous, deep root system. Sorghum leaf blades, similar in appearance to those of its close relative maize, may be up to a meter in length with a width of 10–15 cm. As a drought tolerant, nutrient use efficient, C_4 species, it is highly productive and resilient.

The panicles and grains of the *Sorghum* species can vary widely in shape and size and represent a means for racial classification. Sorghum panicles are made up of perfect flowers, and it is considered a self-pollinated species; however, outcrossing can be as high as 70 % in certain races in particular environments.

Originating in Sudan and Ethiopia, sorghum in nature is a short-day plant and photoperiod sensitive. Elite germplasm, however, has been bred to be photoperiod insensitive with multiple maturity classes based on manipulation of various known maturity genes. Sorghum is predominantly propagated by seed, but some species (*S. halepense* and *S. propinquum*) expand their cover through the production of rhizomes. As such, *Sorghum* species are represented by both annual (maturing in approximately 60–180 days) and perennial types.

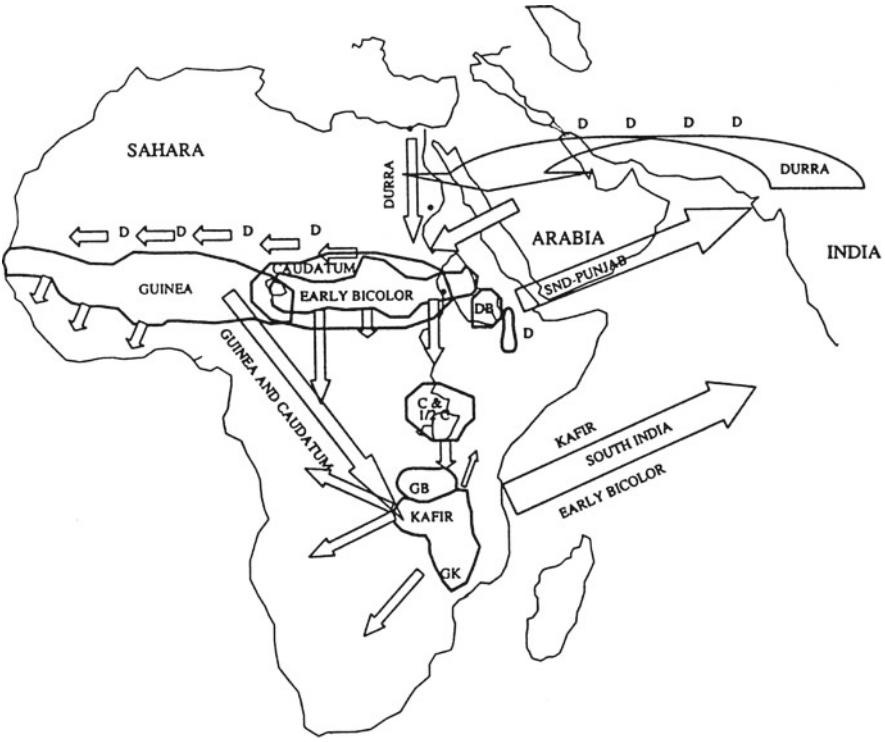

Fig. 2.1 Areas of origin and development for the domesticated races of *bicolor* and possible migration routes

1.2 Evolution and Domestication of the Species

The largest diversity of cultivated and wild sorghum is in Africa (Doggett 1970; de Wet and Harlan 1971; de Wet 1977). The great diversity of *S. bicolor* was created through disruptive selection and by isolation and recombination in the extremely varied habitats of northeast Africa and the movement of peoples carrying the species throughout the continent (Doggett 1970, cited in Miller 1982). There seems no argument that sorghum plants are African in origin, with the earliest known record of sorghum found in 8000 BP charred remains of sorghum at the Nabta Playa in Southern Egypt (Dahlberg and Wasylikowa 1996), but the domestication event(s) may also have taken place elsewhere and more than once.

Based on experiments and on the work of Snowden (1936), Harlan (1995) and his associates confirmed that "all of the races belong to the same biological species and are fully fertile when hybridized." For the cereal sorghums, they identify four wild races and five cultivated races (Harlan and Stemler 1976). The four wild races of *Sorghum bicolor* are arundinaceum, virgatum, aethiopicum, and verticilliflorum. They are now placed in *S. bicolor* subspecies *verticilliflorum*, formerly subspecies *arundinaceum*.

Using de Wet and Rao's interpretation (de Wet and Rao 1986), Doggett (1988) puts the four weedy races under *S. bicolor* subspecies *verticilliflorum* even though de Wet earlier (de Wet 1977) thought they were well-defined ecotypes and should not be given formal taxonomic status.

The cultivated races as presently conceived are (1) bicolor, the primitive type, (2) guinea, (3) kafir, (4) caudatum, and (5) durra. Intermediates that are caused by hybridization of races exhibit characters of both parents. All will also breed with wild species with which they are sympatric. These cultivated races are placed in *S. bicolor* subspecies *bicolor*. Comparison of the distribution of the wild and cultivated races indicates a narrower concentrated band of the cultivated races in the east–west line across and north–south line on the eastern part of the continent than for the wild races. Such a distribution is support for the dispersal of sorghum taxa by migrating peoples across the Sahel-Sudan grasslands and southward from the Nile Valley region along the Great Rift (Murdock 1959; Harlan 1995).

Snowden (1936) took the position that sorghum had separate centers of origins for different types. According to Snowden, wild race aethiopicum gave rise to races durra and bicolor, arundinaceum to guinea, and verticilliflorum to kafir. de Wet and Huckabay (1967) had much the same understanding except they proposed that durras came out of kafirs. Doggett (1965) suggested that the diversity seen in the wild forms might reflect human manipulation and intervention associated with the selection of domesticated types.

Today, bicolor is distributed widely but is nowhere dominant among the African regions with cultivated sorghums. Bicolor, however, is not only widely distributed in Africa but is also apparently ancient in Asia, coastwise from India to Indonesia to China (de Wet and Price 1976). How bicolor migrated out of Africa is unknown, as are the people who were responsible for its diffusion. Cultivated sorghum may have reached China from Indochina by way of the Mekong River or other river valleys. Hawkes (1973) asserts that Semitic speakers from Africa carried their culture to India before 3000 B.C. This may have been one avenue on which sorghum moved off the coast of Africa and into India.

Today, we see the four wild races distributed throughout the African continent, based on their biological traits. The wild race arundinaceum is distributed largely in wet and humid parts of forested central and west Africa along stream banks and in clearings, which are not suitable environments for the cultivated races. The Nile Valley proper is a region of seasonal flooding, with wild grasses such as *S. virgatum* among the first colonizers after the waters recede. It also is found in disturbed riparian habitats in the Sudan. Based on the biology of this wild species, Harlan's group rejects it as the source of the cultivated types. Race aethiopicum is found in the Kassala region of the Sudan, more sparsely to the west along the fringes of the Sahara, and in Ethiopia. Based on the ecology of the wild races, this leaves *S. verticilliflorum*, widely distributed throughout the sorghum-growing areas, including the savanna zone of eastern and southern Africa (although not well represented in Nigeria), and morphologically appropriate as the parent race for *S. bicolor* (Harlan 1992). This position is a change from the position 20 years previously when Harlan and de Wet considered aethiopicum/verticilliflorum as a complex that was the primitive primogeniture of the cultivated races.

Fig. 2.2 Sorghum phenotypic diversity for height (*left*) and panicle (*right*)

The cultivated races also are found in different regions of Africa according to their biological traits and also their histories of distribution (see Fig. 2.2). The guinea race is basically a West African race but also has a distribution in the mountains of eastern Africa that receive high rainfall amounts. In an earlier publication, Harlan and Stemler (1976) considered the guineas to be the oldest of the specialized races because of its relatively wide distribution and diversity. It is better adapted to atmospheric conditions that are more wet and humid than are other domesticated races. Its relatively lax panicle provides for the movement of air among the seeds on individual panicle branches. Other scientists have concluded that as bicolor moved west, it came into contact with wild *S. arundinaceum*, and is now found with it in mixed populations, thus picking up some adaptive genetic material through introgression, and from these the race guinea evolved. All three biotypes can be found in feral populations today, especially across the African savannas. Plants of this race can tolerate up to 5,000 mm of rainfall. Guineas moved into east Africa and humid, foggy southeastern Africa and then were transported, probably from east African ports to the Malabar coast of India (Vishnu-Mittre 1974; Harlan and Stemler 1976; Kimber 2000, Fig. 2.2).

According to Stemler et al. (1975), race caudatum is a later domesticate than bicolor and guinea, having been segregated out of bicolor in the ancestral home territory of that race. Caudatum is associated with peoples speaking the Chari-Nile language group. Groups in areas outside the territory of these peoples use the caudatums; however, such groups have only been within the trade area of the Chari, such as Ethiopia and Cameroon (Stemler et al. 1975). Stemler and her associates concluded that caudatums of the Ethiopian highlands are intrusive, probably acquired by trade. These plants, described as being adapted to harsh conditions, are found most commonly in areas receiving from 250 to 1,300 mm of rain annually. (Stemler et al. 1975, 1977).

De Wet (1978) and Harlan et al. (1976) concluded that race kafir was derived from an early bicolor race, which had been carried east and south from the Savanna belt. Electrophoresis data collected by Schecter and de Wet (1975) suggest that the kafirs are more closely associated with wild race verticilliflorum. The kafir sorghums are very much associated with the Bantu-speaking peoples of eastern and southeastern Africa. It is known that the Bantu entered this part of Africa from the

western forested regions to the north and may have migrated before the guinea race was segregated from the race bicolor. Kafirs may have been taken to the Indian plateau edge ports after discovery of the monsoon wind systems by traders in the western Indian Ocean. This route is different from the route via the northern ports, so it may be a later development than the movement of bicolors to the Indus Valley and Punt, or it may be a second migration into the southern plateau area, having been preceded by bicolor.

The compact panicle and predominantly white seeds of race durra are indications of adaptation to low-rainfall environments with a low risk of grain mold (Mann et al. 1983). The most important grain in Ethiopia, it is found strictly north of the equator in Africa. It is an important type in India and may have been domesticated there (Harlan and Stemler 1976). The name is derived from an Arabic root. Until recently, the durras were almost entirely cultivated by Muslim Africans and Arabic people in Ethiopia. Harlan et al. (1973) reported that the main growers of durra sorghum in Ethiopia are the Muslim Oromo (Gallo), who settled the fertile warm highland almost 500 years ago and have used race durra sorghum as the foundation of their agricultural system. Durras are presently distributed in the mid-altitude highlands of Ethiopia, the Nile Valley of Sudan and Egypt, and in a belt 10–15° north latitude from Ethiopia to Mauritania. They are grown also in the Islamic and Hindu areas of India and Pakistan.

The kaoliangs (galiangs) of China are also thought to be derived from races of bicolor introduced from India to China (Harlan 1995). Alternatively, they may be derived from wild diploid sorghums with which they were compatible (Harlan 1995). At one time they were considered native to Manchuria (Quinby 1974). According to Harlan (1995), the Chinese kaoliangs are more fibrous and have been selected for nongrain uses such as basketry, fencing materials, and house construction materials. Broomcorn sorghum is thought to be part of the story, as are the sorgos such as amber cane. In some regions, the Chinese use the grain in the preparation of a fiery whiskey called mai-tai.

In Southeast Asia and Indonesia, the sorghum are different as well. *S. propinquum* is found in southern China through Thailand, Cambodia, Malaya, and Burma to the Philippines (Burkill 1966). These sorghum are characterized by very large, loose, open panicles and may also have a history different from those of the African-based races (Harlan 1975; Doggett 1988). In fact, recent molecular studies show that *S. propinquum* exhibits approximately an 1.2 % nucleotide difference in coding regions of the genome from *S. bicolor*, suggesting a divergence of 1–2 million years between the two sorghum (Feltus et al. 2004).

1.3 Domestication and Its Effect on Genome and Gene Pool Diversity

Over the past decade, tremendous progress has been made to build the molecular and genomic foundation required to increase our understanding of sorghum diversity in

the genome and gene pool. In complement, these genomic resources currently are being deployed in sorghum improvement efforts across the world. Sorghum represents the first crop genome of African origin to be sequenced (Paterson et al. 2009) and, through coordinated national and international efforts, high-density genetic and physical maps, extensive sets of RFLP and SSR markers, association (Casa et al. 2008) and diversity panels (Deu et al. 2006), nested association mapping populations, etc., are readily available for use for scientific investigations and breeding efforts.

The generation and use of these genomic resources have added to our insights of sorghum domestication and diversity. Future studies will enrich our understanding and provide increasing resolution to quantify and use both wild and domesticated sources of diversity in crop improvement.

At the genomic level, diversity studies have focused on measure of neutral diversity (simple sequence repeats, single nucleotide polymorphisms, etc.), genic diversity, and linkage disequilibrium. In all cases, the domestication process has had a major effect on how much diversity is present and how it is organized in the genome and gene pool.

In recent years, a number of studies of genomic architecture and diversity of sorghum have been undertaken (Casa et al. 2005, 2006, 2008; Hamblin et al. 2004, 2005, 2006; de Alencar Figueiredo et al. 2008). Hamblin and her colleagues (2004) observed an average of one single nucleotide polymorphism (SNP) about every 120 nucleotides of a sample that included a survey of approximately 30 cultivated and wild sorghum accessions and 96 loci representing 29,186 bases of DNA. This is about one-fourth the frequency observed in a comparable sample of maize (Tenaillon et al. 2001). Subsequent studies, across a broader range of sorghum, have yielded similar results (de Alencar Figueiredo et al. 2008).

Studies of linkage disequilibrium in sorghum are of interest because they provide evidence of both equilibrium (e.g., mating system or long-term population structure) and nonequilibrium (e.g., demographic or selective) processes, and also because of their importance in strategies for identifying the genetic basis of complex traits of importance to agriculture. Hamblin and her associates (2005) randomly surveyed six unlinked genomic regions of sorghum and found that patterns of linkage disequilibrium ranged from a few thousand to tens of thousands of bases. An average value of approximately 15 kb suggests that sorghum may be well suited for association studies using a reasonable number of markers. This is in sharp contrast to maize where linkage disequilibrium decays in less than 2 kb in many instances. Moving from studies of genomic to gene pool diversity also has been rapid; this transition likely occurred because of the wealth of molecular tools accessible in sorghum.

Deu and her associates (2006) evaluated a diverse core collection of 210 accessions representing the cultivated races. Two major geographic poles for sorghum evolution and differentiation were established (northern and southern equatorial types). The absence of rare alleles in the southern equatorial accessions (kafir, guinea, and caudatum) fits with the classical view that southern equatorial sorghums evolved later from other African sorghum. Additionally, in support of the historical studies, Deu found that morphological race had a major effect on patterns of genetic diversity.

Additionally, Deu and her colleagues (2006) quantified amount of diversity within race (based on a measure of gene diversity, mean number of alleles, its presence within the defined clusters of accessions) noting that the greatest levels were found with the races bicolor and guinea. Kafir represented the race with the most limited representation of diversity, and this finding likely reflects the classical view of this race and its recent origin and restricted geographic distribution. The assessment of kafir's limited diversity (based on restriction fragment length polymorphisms) is consistent with a complementary simple sequence repeat analysis of wild and domesticated sorghum accessions conducted by Casa and her colleagues (2005).

Like in other crop species, the process of domestication has reduced the amount of diversity present in the cultivated gene pool of the species. Casa and her colleagues (2005) noted the reduction across loci (as measured by simple sequence repeat analysis) was equivalent to 86 % of the diversity as observed in the wild sorghums. Statistical methods for identifying genomic regions with patterns of variation consistent with selection yielded approximately 10 % of the screened loci possibly under some selection pressure. Interestingly, approximately two-thirds of these loci mapped in or near genomic regions associated with domestication-related QTLs (including seed shattering, seed weight, and rhizomatousness).

However, domestication and crop improvement is not always a "one-way" street representing an ever-present move toward reduced diversity. De Alencar Figueiredo and others (2008) studied diversity relationships in sorghum for six candidate genes associated with grain quality and established that in some cases, the genes *Waxy* and *Amylose extender 1*, novel variation was detected (via positive selection), suggesting that postdomestication mutations had occurred and were seen as desirable. Also, these variants subsequently were conserved and increased in frequency and range under human selection. These recent findings have important implications for genetic diversity conservation and use. With increasing access to and cost-effective use of molecular tools, curators now will be able to establish high-resolution studies of diversity in their collections and in natural settings.

2 Centers of Diversity: Current Status of Sorghum Collections

2.1 Global Status of Sorghum

Investigators continue to collect new and additional land races from isolated farmers' fields (Benor and Sisay 2003), and various collections have been assembled world-wide that represents much of the genetic diversity that can be found in sorghum. Major collections from the site of origin of sorghum exist for Ethiopia and Sudan, while other collections representing the major races and working groups of sorghum have been compiled in areas such as Mali, South Africa, India, and China. Internationally, two collections have worked to bring most, if not, all of these various collections together; the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) located in Andhra Pradesh, India and the National Center for Genetic Resources Preservation, located in Fort Collins, Colorado, USA.

Table 2.1 Major sorghum collections from various countries from around the world

Country	Institute	No. of accessions	% of total holdings (194,250 acc.)
USA	USDA-ARS-PGRCU	43,104	22.2
Global	ICRISAT	36,774	18.9
India	NBPGR	18,853	9.7
China	CAAS	18,250	9.4
Ethiopia	IBC	9,772	5.0
Brazil	EMBRAPA	8,017	4.1
Russia	VIR	7,335	3.8
Zimbabwe	NPGRC	7,009	3.6
Australia	DPI	5,403	2.8
Sudan	PGRU-ARC	4,191	2.2
Mali	IER	2,975	1.5
France	CIRAD	2,690	1.4
Kenya	NGBK	1,320	0.7
Zambia	NPGRC	1,005	0.5
South Africa	NPGRC	428	0.2
Malawi	NPGRC	401	0.2
Nigeria	NCGRB	159	0.1
Serbia	Inst. Field and Vegetable crops	152	0.1
Global	ILRI	52	0.0
	Total 19 institutes	167,890	86.0

(From the "Strategy for the Global Ex Situ Conservation of Sorghum Genetic Diversity")

In 2006, the Global Crop Diversity Trust initiated a review of the sorghum collections from the Germplasm Holding Database maintained by Bioversity International (formerly known as IPGRI). This became a major attempt to catalogue collections from around the world and review the condition of those collections. One hundred twenty-two collections were identified and based on input from various experts, a global survey was sent to 57 institutes to gather information on collection numbers, state of preservation, regeneration procedures, and other information regarding each collection. Nineteen institutes responded and from this the beginnings of a world-wide inventory began to form. From this initial review, the Global Crop Diversity Trust brought together sorghum experts from around the world in 2007 to formulate the "Strategy for the Global Ex Situ Conservation of Sorghum Genetic Diversity," which met at ICRISAT (see report at http://www.croptrust.org/documents/cropstrategies/sorghum.pdf, verified May 5, 2010). Major collections are outlined in Table 2.1.

It became clear from this undertaking that many of these collections represent duplications of various sorghum accessions. One of the major challenges identified by this working group was the lack of relevant passport data available on these collections which could assist in grouping duplicate accessions and developing a more accurate picture of the true collection status of sorghum. Another major concern was the lack of wild relative species within these collections. Only 159 wild relative species were identified in the 19 collections that sent back their surveys, and it is questionable as to how much of that collection has been correctly identified and classified.

2.2 Major Collections

2.2.1 The Americas

The single largest collection of sorghum germplasm resides in the USA, which began collecting various forms of sorghum as early as 1757, when Benjamin Franklin mentioned it in a letter to a Mr. Ward. He was interested in its unique panicle formation, used for making brooms and had brought seed back from Europe to the USA to increase and to share with friends. The Philadelphia Agricultural Society mentioned Guinea Corn in 1810 (Quinby 1974) and references to sorghums such as sorgo, Chinese amber cane, white and brown durras, milo, feterita, and hegari can be found in various publications between 1853 and 1908 (Doggett 1988). Formal collections and distributions by the United States Department of Agriculture began around 1905 when it bought, increased and distributed Dwarf Yellow Milo; further introductions and research began at the Texas Agricultural Experiment Station at Chillicothe, Texas (Quinby 1974). Prior to the introduction of hybrids in 1958, approximately 13,611 accessions of sorghum had been introduced into the USA; however, serious curation of the crop did not take place until the early 1980s. Since then a total of 31,163 accessions have been added to the collection for a total of 44,774 accessions (Table 2.1). Several groups have reviewed the status of the USA collection (Duncan, Bramel-Cox, and Miller 1991; Dahlberg and Spinks 1995). The second largest collection is housed with EMBRAPA in Brazil and is used extensively in breeding programs used to search for abiotic and biotic source of resistance to things such as anthracnose and acid-soil tolerance, two major issues facing sorghum producers in this country. Both Mexico and Argentina maintain collections of over 3,000 accessions for use in their national breeding programs and for research purposes.

2.2.2 Africa

Several collections exist in Africa, the site of sorghum domestication and early distribution (Mann et al. 1983). The Ethiopian and Sudanese collections are two of the most important worldwide, since these two regions are considered the primary sites of sorghum domestication. The discovery of the 8000 BP charred remains of sorghum at the Nabta Playa in Southern Egypt (site E-75-6) were identified by de Wet and Harlan (sorghum taxonomy according to de Wet 1978) as wild *Sorghum bicolor* (L.) Moench. subsp. *arundinaceum* (Desv.) based on spikelet morphology and the small size and shape of the grains (Dahlberg and Wasylikowa 1996). The Jimma Agricultural Technical School began to centralize collections in Ethiopia between 1958 and 1960, and the collection was grown out and characterized in an Experiment Station Bulletin entitled “The Cultivated Sorghums of Ethiopia.” This activity was taken over by the Ethiopian Sorghum Improvement Project and continued collections have taken place since the 1970s. Sudan collected a large collection of

landraces in the 1950s which was maintained by the Tozi Research station. This collection was turned over to the Rockefeller Foundation Project in India (Rosenow and Dahlberg 2000). In 1992, collections with known Sudanese origins were grown in Wad Medani for increase, characterization, and distribution. This collection was shared with the USA and 3,182 of these accessions were grown out in St. Croix, USA Virgin Islands in 1993 for verification, seed increase and characterization. The Sudanese collection may be one of the most fully characterized collections in the world to date and contains representatives of most of the major working groups (Dahlberg and Spinks 1995; Dahlberg and Madera-Torres 1997; Rosenow and Dahlberg 2000; Dahlberg, Burke, and Rosenow 2004, Grenier, Bramel, Dahlberg, El-Ahmadi, Mahmoud, Peterson, Rosenow, and Ejeta 2004). Uganda also maintains approximately 2,600 accessions for use in its research programs.

The largest collection outside of Eastern Africa is located with NPGRC in Zimbabwe, with over 7,000 accessions. Western Africa is represented by collections from Mali, Niger, Nigeria, and Burkina Faso, but their numbers are more difficult to pin down. The Mali collection was shared with the USA and a joint project involving the USA, ICRISAT, CIRAD, ORSTOM, and Mali was undertaken in 1997, to grow out, increase, characterize, and distribute the collection in Mali. The collection was also regrown in St. Croix and data sets from both countries were incorporated into the GRIN database. Other smaller collections exist in Africa and represent some of the unique diversity that is found in the world collection.

2.2.3 Asia

The second largest collection worldwide is held in trust at ICRISAT. K.O. Richie transported a Rockefeller Foundation collection of sorghum from Mexico to India in 1957, thus starting the international collection of sorghum. The collection was a cooperative effort between the Rockefeller Foundation and the All-India Sorghum Improvement Program. Comprised of various collections from around the world, this collection began an intensive effort to collect indigenous sorghums from places not fully represented in those early collections (House 1980). International Sorghum numbers were given to the collection, and it grew to its current status of approximately 36,774 accessions. Both China (Qingshan and Dahlberg 2001) and India have collections of over 18,000 accessions; however, the Chinese collection is not well represented in either the ICRISAT or US collection. Much of the diversity of the Indian accessions has been collected in the International collection. Japan and the Philippines have small research collections.

2.2.4 Other Regions

Various other collections exist worldwide; however, it is difficult to know how much of these collections represent unique accessions of sorghum or duplicates of germplasm from various regions from around the world. Russia has approximately

7,000 accessions of sorghum, while Australia reported a collection size of 5,403. Many of the other collections are less than 2,000 and most have been obtained for research purposes in national sorghum improvement programs. With over 160,000 accessions reported in various collections, it is difficult without better phenotypic and genotypic information to fully understand how much diversity of sorghum has been duplicated in the world collections.

3 Assessing Useful Diversity, Breeding, and Race and Working Groups

3.1 Assessing Phenotypic Diversity

The international sorghum community has shown a tremendous amount of interest in evaluating germplasm collections. One of the many difficulties that the working group identified as they worked on the Strategy for the Global Ex Situ Conservation of Sorghum Genetic Diversity was how to bring those data points together in a useful database that could be available for worldwide use.

Several issues were identified that limit the overall utility of the various collections. Passport data was available in some electronic form, but different nomenclature use by the various institutes or collecting agencies made comparison of collections problematic, especially as to the when, where, and how accessions were collected. Different databases, ratings, evaluation techniques, and other issues have made a simple combining of the various data points difficult. Few of the collections have longitude and latitude designations, making spatial evaluation of the collections almost impossible. Many of the collections were done pre-GIS technology and relied on local names and/or villages as collection points. These same concerns are also true for both characterization and evaluation data.

It is clear that there is tremendous diversity within the sorghum collections. The USA has clearly shown this in its descriptors and through various evaluations that they have undertaken on their national collection; however approximately 33 % of the collection has extensive phenotypic descriptor data, while roughly 60 % of the collection has race and working designation, and 40–60 % of the collection has agronomic characteristics (Table 2.2). The collection has been evaluated for photoperiod response in a temperate zone (Table 2.3).

The ICRISAT collection's database contains 24 data points on each of its accessions. Most of the evaluation points reflect agronomic phenotypic characterization and is available from the SINGER database. The Chinese collection has approximately 50 % of its collection characterized, but most of that is not available electronically (see Qingshan and Dahlberg 2001). CIRAD has significant evaluation data, but until datasets can be electronically merged and validated, accessing the phenotypic value of the various collections will be difficult.

Table 2.2 Phenotypic characteristics and evaluation data points taken on various accessions within the US National Sorghum Collection (source GRIN database)

Descriptor	Distinct accessions	Total observations
Chemical		
Acid detergent fiber %	2882	2,914
Brix measurements	1229	1,260
Crude protein %	2882	2,914
In vitro dry matter digestibility	2881	2,914
Fat %	2882	2,913
Phosphorous %	2882	2,914
Starch %	2882	2,914
Sucrose %	1198	1,229
Cytological		
Normal ploidy level	49	49
Disease resistance		
Colletotrichum graminicola	15670	16,399
Peronosclerospora sorghi	4674	6,214
Peronosclerospora sorghi P3	3937	3,938
Claviceps africana	2022	2,022
Cercospora sorghi ELL. and Ex.	306	306
Cercospora fusimaculans	1437	1,470
Helminthosporium turcicum	340	340
Puccinia purpurea	15819	17,402
Sugarcane mosaic virus	427	427
Sorghum yellow banding virus	210	210
Gloeocercospora sorghi	1437	1,470
Growth		
Height uniformity of plant	15651	16,149
Plant height (to top of panicle)	20195	21,577
Seedling vigor	3984	3,987
Insect resistance		
Pseudaletia unipuncta (Fall armyworm)	8940	8,942
Taxoptera graminum (Greenbug)	15990	16,035
Sipha flava (Yellow sugarcane aphid)	5564	5,564
Restorer		
B/R line reaction to A1, A2, and A3	620	1,828
Morphological screening		
% Glume covering of kernel	14655	15,123
Type of awns at maturity	15615	16,101
# of basal tillers per plant	14690	15,158
Panicle branch angle	14637	15,106
Color of endosperm	14540	15,003
Texture of endosperm	14537	15,000
Type of endosperm	14547	15,008
Color of glume	14659	15,128
Glume pubescence	14657	15,127
Inflorescence exsertion	14674	15,139
Phenotypic seed color	18745	19,135

(continued)

Table 2.2 (continued)

Descriptor	Distinct accessions	Total observations
Plumpness of kernel	14532	14,992
Shape of kernel	14540	15,003
% lodging	965	965
Mesocarp thickness	14651	15,119
Leaf midrib color	19081	19,770
Nodal tillering	14680	15,147
Compactness of panicle	15610	16,109
Panicle erectness	15637	16,132
Length of panicle	14641	15,105
Shape of panicle	14663	15,132
Color of pericarp (genetic)	14362	14,791
Plant color	14743	15,210
Form of seed (single, twin)	14642	15,111
Seed shattering	14651	15,121
Spreader	13150	13,479
Seed sprouting tendency	14687	15,155
Juicy/dry midrib	19054	20,401
Waxiness of the stem	14664	15,131
Presence/absence of testa	14713	15,182
Transverse wrinkle	14201	14,666
Other		
Cold tolerant population	173	174
Core subset	2438	2,438
Images present in GRIN	6880	7,089
Sorghum association panel	386	387
Phenology		
Flowering rating 65 and 90 days after planting	32680	32,718
Long day anthesis rating	739	742
Photoperiod sensitivity rating	1533	1,533
Short day anthesis rating	12437	12,763
Production		
Primary plant usage	1390	1,390
Yield potential (1 = high, 5 = low)	14665	15,130
Overall plant desirability ratings	14729	15,196
Quality		
Grain weathering	14657	15,126
Metabolizable energy (swine)	2882	2,914
Net energy for gain (cattle)	2882	2,914
Net energy for lactation (cattle)	2882	2,914
Net energy for maintenance (cattle)	2882	2,914
% Total digestable nutrients	2882	2,914
Stress		
Aluminum toxicity tolerance	10332	10,384
Manganese toxicity tolerance	7302	7,339
Taxonomy		
Race	22319	23,011
Working group designation	14805	15,262
Total number of observations		7,88,302

Table 2.3 Photoperiod rating of US National Sorghum Collection (source GRIN database)

Definition	Number of accessions
Very early (50 days or less)	197
Early (50–60 days)	1,473
Medium (60–70 days)	3,796
Late (75–90 days)	3,498
Very late (90+ days)	925
Photoperiod sensitive	22,762
Mixed flowering rating	66
Total number of accession screened	32,717

3.2 Breeding Collections

Little work has been done on establishing a robust acquisition system for breeding collections as sorghum breeders retire or pass away. Several collections have been either discarded or archived with little recorded evidence of their existence. This is also true within the private industry as collections have been either neglected or destroyed when a program has ended or a breeder has retired. Several important collections have been noted in research articles and other publications, but are no longer available for distribution. The USDA-ARS tried reviving Dr. Keith Schertz's genetic stocks after he had passed away; however, little of the collection was maintained properly and only a few of his genetic stocks were recovered and placed into GRIN. No concerted effort worldwide has been made to preserve these potentially valuable collections.

Probably one of the most useful breeding collections has been the Sorghum Conversion Program, which has recently been relaunched through the support of the United Sorghum Checkoff Program and MMR Genetics. The program was first described by Stephens, Miller, and Rosenow in 1967, though it was initiated in 1963 with the first planting taking place in Mayagüez, Puerto Rico. They recognized the need to convert photoperiod sensitive sorghums, to combine height, photoperiod insensitive germplasm that would flower in temperate regions. This was partly driven by the need to expand the germplasm base that was currently in use in the USA. As noted earlier, sorghum is unique in that it is grown from sea level to elevations of 2,500 m or more, in both extremely wet and dry environments, and on both poor and rich soils. This has created a tremendously diverse and genetically rich germplasm base that has been mostly ignored because of its photoperiod sensitivity and the difficulty of using these germplasm sources in temperate zones.

As these accessions were converted, many were screened for various diseases and insects and several accessions were identified containing key genes of valuable agronomic characteristics (Duncan et al. 1991). Currently, 702 converted lines have been released. These releases ranged from exotic germplasm that was initially placed into the program for diversity, to agronomically important, high yielding white, tan food grade sorghums (see discussion by Rosenow and Dahlberg 2000). In 2009, the sorghum conversion program was reinitiated and is exploiting the

sorghum sequence to rapidly convert sorghum exotic germplasm sources. Typically, the conversion program took between 8 and 12 years to fully convert a photoperiod sensitive sorghum to a converted insensitive line. Using DarT technology, and markers for both the Ma_1 gene, which is highly linked to photoperiodism, and several of the dwarfing genes, the program hopes to achieve adequate conversion within 2–3 years. This will allow for a much more rapid distribution of needed exotic germplasm to both private and public breeding programs.

3.3 Working Groups

The cultivated sorghums fall within the subgenera *Sorghum* (see discussion by Dahlberg 2000). Snowden in 1936 proposed 31 races of sorghum, with varieties within each race, which made classification of sorghum quite difficult. As noted previously, Harlan and de Wet proposed a simplified classification scheme for the cultivated sorghum in 1972, which delineated five major races and ten intermediate races, which is used by many plant breeders today to classify sorghum; however, this simplified classification did not adequately explain the variation found within sorghum. Murty and Govil (1967) proposed a set of Working Groups to try to address this. Harlan (personnel communication) indicated that he and de Wet had wanted to integrate their proposed scheme with that of Murty and Govil to create a system for classification that was both easy to use and explicative of the diversity within the crop. Dahlberg (2000) proposed an integrated system of sorghum classification with this goal.

Little work has been done to evaluate and relate important genetics with various races and working groups, though it is clear that such an endeavor would be useful to sorghum researchers. Working groups such as Zerazeras contain useful sources of tan plant and white seed for use in food systems, while kafirs offer good sources of yield. Some work has been done on heterotic pools, with caudatum and kafir combinations being the most widely used. However, the work by Yang and colleagues (Yang et al. 1996) clearly indicates that Chinese germplasm may offer new sources of heterosis that have not been exploited within sorghum. Other potential sources of new heterotic pools between durras and bicolors offer some yield potential that would be new to sorghum as well (personal observations by author of diversity within the Malian collection). New genomic tools may offer the potential to exploit these pools in the future as more of the world collection becomes fingerprinted and evaluated for genetic diversity.

4 The Future

As competition for global water resources becomes more intense in the twenty-first century, sorghum will increase in importance as a source of food, feed, fiber, and fuel. However, key goals must be accomplished to support the wider use of sorghum.

First, wild and weedy relatives of cultivated sorghum must be effectively collected and conserved. Current collections across the world are void of the exceptional range and diversity of the genus. Particularly wild and weedy materials from Africa and Asia merit inclusion into international and national collections. Second, entries in the collection require more detailed characterization and evaluation if elite germplasm is to be effectively developed. In particular, photoperiod sensitive materials require conversion for appropriate agronomic evaluation. An expanded effort for sorghum conversion also should be a priority.

In complement with the agronomic benefits of increased sorghum production for multiple end-products, sorghum will increase in value for studies of comparative genomics in cereals and grasses (Paterson et al. 2009). Sorghum, because of its smaller and simpler genome, its wealth of readily accessible genomic tools and information, and its breadth of diversity for abiotic and biotic stress resistances will become a cornerstone for integrated studies linking gene pool diversity and crop improvement in the future.

Acknowledgments The authors would like to gratefully acknowledge Elizabeth Herron in the Office of Research Development at the University of South Carolina for her assistance with editing and fact-checking.

References

Benor S, Sisay L (2003) Folk domestication of sorghum (Sorghum bicolour (L) Moench) land races and its ethnobotanical implications: a case study in northeastern Ethiopia. Ethnobiologia 3:2941

Burkill IH (1966) A dictionary of the economic products of the Malay Peninsula, vol 1, 2nd edn. A.H. Government of Malaysia and Singapore, Kuala Lumpur, Malaysia

Casa AM, Mitchell SE, Hamblin MT, Sun H, Bowers JE, Paterson AH, Aquadro CF, Kresovich S (2005) Diversity and selection in sorghum: simultaneous analyses using simple sequence repeats (SSRs). Theor Appl Genet 111:23–30

Casa AM, Mitchell SE, Jensen JD, Hamblin MT, Paterson AH, Aquadro CF, Kresovich S (2006) Evidence for a selective sweep on Chromosome 1 of cultivated sorghum. Crop Sci 46:S27–S40

Casa AM, Pressoir G, Brown P, Mitchell SE, Rooney WL, Tuinstra MR, Franks CD, Kresovich S (2008) Community resources and strategies for association mapping in sorghum. Crop Sci 48:30–40

Dahlberg JA (2000) Classification and characterization of sorghum. In: Smith CW, Frederiksen RA (eds) Sorghum: origin, history, technology, and production. Wiley, New York, pp 99–130

Dahlberg JA, Burke JJ, Rosenow DT (2004) Development of a sorghum core collection: refinement and evaluation of a subset from Sudan. Econ Bot 58:556–567

Dahlberg JA, Madera-Torres P (1997) Restorer reaction in A_1 (ATx623), A_2 (A_2Tx632), and A_3 (A_3SC103) cytoplasms to selected accessions from the Sudan sorghum collection. ISMN 38:43–58

Dahlberg JA, Spinks M (1995) Current status of the U. S. sorghum germplasm collection. Sorghum Newsl 36:4–11

Dahlberg JA, Wasylikowa K (1996) Image and statistical analyses of early sorghum remains (8000 B. P.) from the Nabta Playa archaeological site in the Western Desert, southern Egypt. Veg Hist Archaeobot 5:293–299

De Alencar Figueirdeo LF, Calatayud C, Dupuits C, Billot C, Rami JF, Brunel D, Perrier X, Courtois B, Deu M, Glaszmann JC (2008) Phylogeographic evidence of crop neodiversity in Sorghum. Genetics 179:997–1008

Deu M, Rattunde F, Chantereu JA (2006) A global view of genetic diversity in cultivated sorghums using a core collection. Genome 49:168–180

de Wet JMJ (1977) Domestication of African cereals. Afr Econ Hist 3:15–32

de Wet JMJ (1978) Systematics and evolution of sorghum sect. Sorghum (gramineae). Am J Bot 65:477–484

de Wet JMJ, Harlan JR (1971) The origin and domestication of *Sorghum bicolor*. Econ Bot 25:128–135

de Wet JMJ, Huckabay JP (1967) The origin of sorghum bicolor. II. Distribution and domestication. Evolution 21:787–802

de Wet JMJ, Price EG (1976) Plant domestication and indigenous African agriculture. In: Harlan JR, De Wet JMJ, Stemler A (eds) Origins of African plant domestication. Mouton, The Hague, pp 453–464

de Wet JMJ, Rao KEP (1986) Wild sorghums and their significance in crop improvement. In: Proceedings of sorghum conference, 20–23 August 1986, Shenyang, China. Cited in J. Dahlberg (2000)

Doggett H (1965) Disruptive section in crop development. Nature 4981:279–280

Doggett H (1970) *Sorghum* (Tropical Agriculture Series). Longmans, London

Doggett H (1988) Sorghum. Wiley, New York

Duncan RR, Bramel-Cox PJ, Miller FR (1991) Contributions of introduced sorghum germplasm to hybrid development in the USA. In: Shands HL, Wiesner LE (eds) Use of plant introductions in cultivar development: part 1. CSSA Spec. Publ. No. 17. CSSA, Madison, WI, pp 69-102

Feltus FA, Wan J, Schulze SR, Estill JC, Jiang N, Paterson AH (2004) A SNP resource for rice genetics and breeding based on subspecies *indica* and *japonica* genome alignments. Genome Res 14:1812–1819

Grenier C, Bramel PJ, Dahlberg JA, El-Ahmadi A, Mahmoud M, Peterson GC, Rosenow DT, Ejeta G (2004) Sorghums of the Sudan: analysis of regional diversity and distribution. Genet Resour Crop Evolut 51:489–500

Hamblin MT, Casa AM, Sun H, Murray SC, Paterson AH, Aquadro CF, Kresovich S (2006) Challenges of detecting directional selection after a bottleneck: lessons from *Sorghum bicolor*. Genetics 173:953–964

Hamblin MT, Mitchell SE, White GM, Gallego J, Kukatla R, Wing RA, Paterson AH, Kresovich S (2004) Comparative population genetics of the panicoid grasses: sequence polymorphism, linkage disequilibrium, and selection in a diverse sample of *Sorghum bicolor*. Genetics 167:471–483

Hamblin MT, Salas Fernandez MG, Casa AM, Mitchell SE, Paterson AH, Kresovich S (2005) Equilibrium processes cannot explain high levels of short- and medium-range linkage disequilibrium in the domesticated grass *Sorghum bicolor*. Genetics 171:1247–1256

Harlan JR (1975) Crops and man. American Society of Agronomy, Madison, WI

Harlan JR (1992) Indigenous African agriculture. In: Cowan CW, Watson PJ (eds) The origins of agriculture: an international perspective. Smithsonian Institution Press, Washington, DC, pp 59–70

Harlan JR (1995) The living fields: our agricultural heritage. Cambridge University Press, Cambridge

Harlan JR, de Wet JMJ (1972) A simplified classification of cultivated sorghum. Crop Sci 12:172–176

Harlan JR, de Wet JMJ, Price EG (1973) Comparative evolution of cereals. Evolution 27:311–351

Harlan JR, de Wet JMJ, Stemler ABL (eds) (1976) Origins of African plant domestication. Mouton, The Hague

Harlan JR, Stemler A (1976) The races of Sorghum in Africa. In: Harlan JR, de Wet JMJ, Stemler A (eds) Origins of African plant domestication. Mouton, The Hague, pp 465–478

Hawkes J (1973) The First Great Civilizations. Penguin Books, Hammondsworth, Middlesex
House LR (1980) A guide to sorghum breeding. International Crops Research Institute for the Semi-Arid Tropics, Patancheru
Kimber C (2000) Origins of domesticated sorghum and its early diffusion to India and China. In: Smith CW, Frederickson RA (eds) Sorghum: origin, history, technology, and production. Wiley, New York, pp 3–98
Mann JA, Kimber CT, Miller FR (1983) The origin and early cultivation of sorghums in Africa. Texas Agric. Exp. Stn. Bull. 1454
Miller FR (1982) Genetic and environmental response characteristics of sorghum. International Crops Research Institute for the Semi-Arid Tropics Symposium on Sorghum, 2–7 November 1981, Patancheru, India, pp 393–446
Murdock G (1959) Africa: its peoples and their culture history. McGraw Hill, New York
Murty BR, Govil GN (1967) Description of 70 groups in genus Sorghum based on a modified Snowden's classification. Indian J Genet Plant Breed 27:75–90
Qingshan L, Dahlberg JA (2001) Chinese sorghum genetic resources. Econ Bot 55:401–425
Quinby JR (1974) Sorghum improvement and the genetics of growth. Texas A&M University Press, College Station, TX
Paterson A, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellston U, Mitros T, Poliakov A, Schmutz J, Spannag M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, ur-Rahman M, Ware D, Westhoff P, Mayer KFX, Messing J, Rokhsar DS (2009) The Sorghum bicolor genome and the diversification of grasses. Nature 457:551–556
Rosenow DT, Dahlberg JA (2000) Collection, conversion, and utilization of sorghum. In: Smith CW, Frederiksen RA (eds) Sorghum: origin, history, technology, and production. Wiley, New York, pp 309–328
Schecter Y, de Wet JMJ (1975) Comparative electrophoresis and isozyme analysis of seed proteins from cultivated races of sorghum. Am J Bot 62:254–261
Snowden JD (1936) The cultivated races of sorghum. Adlard & Sons, Ltd., London
Stemler ABL, Harlan JR, de Wet JMJ (1975) Evolutionary history of cultivated sorghums (*Sorghum bicolor* [Linn.] Moench) of Ethiopia. Bull Torrey Bot Club 102:325–333
Stemler ABL, Harlan JR, de Wet JMJ (1977) The sorghums of Ethiopia. Econ Bot 31:446–450
Stephens JC, Miller FR, Rosenow DT (1967) Conversion of alien sorghums to early combine genotypes. Crop Sci 7:396
Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). Proc Natl Acad Sci U S A 98:9161–9166
Vishnu-Mittre A (1974) The beginning of agriculture, paleobotanical evidence in India. In: Hutchinson JB (ed) Evolutionary studies in world crops: diversity and change in the Indian subcontinent. Cambridge University Press, Cambridge, pp 3–30
Yang W, de Oliveira AC, Godwin I, Schertz K, Bennetzen JL (1996) Comparison of DNA marker technologies in characterizing plant genome diversity: variability in Chinese Sorghums. Crop Sci 36:1669–1676

Chapter 3
The Gene Pool of *Saccharum* Species and Their Improvement

Andrew H. Paterson, Paul H. Moore, and Tom L. Tew

Abstract Current taxonomy divides sugarcane into six species, two of which are wild and always recognized (*Saccharum spontaneum* L. and *Saccharum robustum* Brandes and Jewiet ex Grassl). The other species are cultivated and classified variously. Of the four domesticated species of *Saccharum*, *S. officinarum* L. was the first named and is the primary species for production of sugar. Recent genomic data for evaluating genetic diversity within *Saccharum* suggests relationships among accessions that may ultimately produce a definitive classification of the species. Sugarcane breeders have long realized that germplasm diversity is essential for sustained crop improvement, with accessions from at least 31 separate expeditions deposited in the two world collections as genetic reservoirs. Cultivated sugarcanes of today are complex interspecific hybrids primarily between *Saccharum officinarum*, known as the noble cane, and *Saccharum spontaneum,* with contributions from *S. robustum*, *S. sinense*, *S. barberi*, and related grass genera such as *Miscanthus*, *Narenga*, and *Erianthus.* Sugarcane has long been recognized as one of the world's most efficient crops in converting solar energy into chemical energy harvestable as biomass, and is of growing interest as a biofactory for production of fossil fuel alternatives and other high-value bioproducts.

A.H. Paterson (✉)
Plant Genome Mapping Laboratory, Departments of Crop and Soil Science, Plant Biology, and Genetics, University of Georgia, Athens, GA, USA
e-mail: paterson@plantbio.uga.edu; paterson@uga.edu

P.H. Moore
Hawaii Agriculture Research Center, Kunia, HI, USA

T.L. Tew
USDA, ARS, Houma, LA, USA

A.H. Paterson (ed.), *Genomics of the Saccharinae*, Plant Genetics and Genomics: Crops and Models 11, DOI 10.1007/978-1-4419-5947-8_3,

Keywords Sugar cane • Interspecific hybrid • Polyploid • Aneuploid • Domestication • Nobilization • Energy cane • Biofactory

1 Singular Properties of the Species (or Its Members)

1.1 Taxonomy

Sugarcane is the common name given to a group of cultivated, sucrose-storing, large tropical grasses that have been classified variously depending on the criteria employed and the taxonomic convention of the time. The original classification of cultivated sugarcane as *Saccharum officinarum* L. by Linnaeus in 1753 established the genus *Saccharum* L. for sugarcane. However, the genus *Saccharum* was expanded to include many species that, with the exception of inflorescence and floral morphologies, have little in common with sugarcane. For example, a search of the Integrated Taxonomic Information System (ITIS) database http://www.itis.gov lists 23 species of *Saccharum* of which 7 are no longer accepted while only 5 are listed as a species of wild or domesticated sugarcane. The more extensive Kew GrassBase database http://www.kew.org/data/grasses-db.html lists 37 species of *Saccharum* of which 4 are currently accepted as species of sugarcane. Current sugarcane literature recognizes six species of *Saccharum*, two of which are wild: *Saccharum spontaneum* L. and *Saccharum robustum* Brandes and Jewiet ex Grassl. The other species are cultivated and recognized variously. This review is restricted to only those *Saccharum* species generally regarded as sugarcane.

Sugarcane species are members of the subtribe Saccharinae, tribe Andropogoneae, of the grass family, Poaceae or Gramineae. The grass family is very large, including approximately 10,000 species classified into 600–700 genera. The tribe Andropogoneae contains 85 genera and 960 species (e.g., Clayton and Renvoize 1986 cited by Spangler et al. 1999), many of which have high economic value including the C4 crops sugarcane (*S. officinarum* L.), sorghum (*Sorghum bicolor* L. Moench), and maize (*Zea mays* L.). *Miscanthus* Anderss. is an additional C4 grass showing considerable potential as a germplasm source for cold tolerance of C4 photosynthesis (Clifton-Brown and Lewandowski 2000) and as a as a biomass crop for renewable energy production and raw material for the cellulose and paper industries (Bullard et al. 1995; Dohleman et al. 2009). Classical taxonomy based on cytological and morphological characters has been used to describe probable evolutionary relationships within Saccharinae and recently molecular data has allowed for more definitive relationships. Spangler et al. (1999) used chloroplast DNA markers to show broad relationships among the Andropogoneae and probable polyphyletic origins of *Sorghum*, *Miscanthus*, and *Saccharum* (Fig. 3.1). Hodkinson et al. (2002) used DNA sequences of a nuclear ribosomal gene and two plastid sequences to confirm the polyphyletic origin of *Miscanthus* and *Saccharum* and to distinguish between them. The other members of the subtribe Saccharinae could not be completely resolved with such limited data.

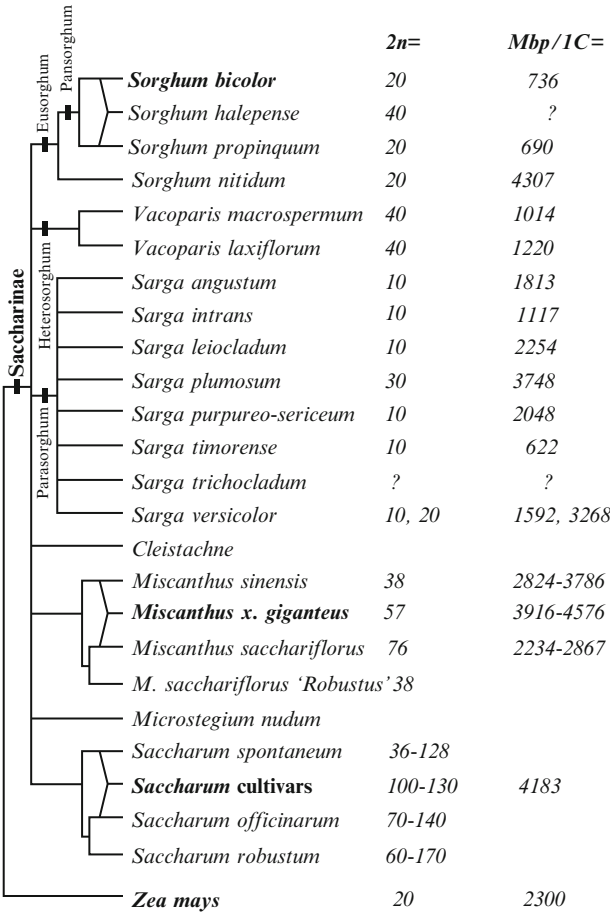

Fig. 3.1 Saccharinae clade (relevant members). Phylogenetic interpretation from Spangler et al. (1999); Genome size estimates from Arumuganathan and Earle (1991), Bennett and Leitch (2003), Price et al. (2005) and direct measurements (*S. propinquum*, *Miscanthus* spp.)

The present compilation of *Saccharum* species has a volatile history. The first edition of "Species Plantarum" listed two species of *Saccharum*: *S. officinarum* and *S. spicatum.* Subsequent taxonomic treatments up to the time of revision by Jeswiet in 1927 listed up to 22 species of *Saccharum* (discussed in Irvine 1999 and Amalrag and Balasundaram 2005, and presented as Table 8 in Daniels and Roach 1987). Jeswiet (1927) reassigned many of those added species to different genera, primarily *Erianthus*, and went on to describe four natural groups consisting of *S. spontaneum* L., *S. sinense* (Roxb.) Jesw., *S. barberi*, Jesw., and *S. officinarum* L. to be included in *Saccharum* L. Subsequently, two forms of *Saccharum* were discovered in New Guinea (*S. robustum* Brandes and Jeswiet ex Grassl. (Grassl 1946) and *S. edule* Hassk. (Whalen 1991)) and added to the genus to bring the number of widely recognized species of sugarcane to six. Two of the species (*S. spontaneum* and *S. robustum*) are

wild accessions. The remaining four cultivated forms (*S. barberi*, *S. edule*, *S. officinarum* and *S. sinense*) have been accorded the status of species, but since they do not survive in the wild there is an increasing trend to consider these as cultigens derived from *S. officinarum* (Irvine 1999, Grivet et al. 2006).

In addition to listing many *Saccharum* that are not sugarcane in the strictest sense, the delineation of these species has been complicated by hybridizations, both natural and man-made, among themselves and with related species. Many members of the interbreeding group have different genome structure which produces intermediate forms, some of which have totally new genome structure due to different types of chromosome transmission. Wide hybridization has resulted in a mixture of sugarcane euploids and aneuploids. Layered on top of this genetic complexity is the selection pressures applied by nature and man to drive different population structures. Cultivars of sugarcane are hybrids of different species of the genus *Saccharum*, and may include germplasm from the nine related genera *Imperata*, *Eriochrysis*, *Eccolipus*, *Spodiopogon*, *Miscanthidium*, *Erianthus* sect. *Ripidium*, *Miscanthus*, *Sclerostachya*, and *Narenga*, which are included in the subtribe Saccharinae (Clayton 1972a, b).

Mukherjee (1954) revised the genus *Saccharum* based on phytogeographical data, morphology, cytology, and breeding evidence to combine *Saccharum* with three other Saccharinae genera (*Erianthus* sect. *Ripidium*, *Sclerostachya*, and *Narenga*) into an informal taxonomic group he called the "Saccharum complex." Later, Daniels et al. (1975) added the genus *Miscanthus* to the Saccharum complex since it was considered that *Miscanthus* also contributed to the origin of *Saccharum*. Although the Saccharum complex concept has proven useful in guiding sugarcane breeders towards utilizing the species within it as part of the gene pool available for sugarcane improvement, recent molecular data raise serious doubts about some of the earlier proposed origins and genetic relationships of the *Saccharum* species (Irvine 1999, Grivet et al. 2006).

The morphological differences among the members of the Saccharum complex are mostly related to floral characters, but also include some vegetative structure characters such as the number of rows of nodal root primordia, axillary bud development, and presence or absence of a leaf dewlap (Table 3.1). A significant difference between members of the Saccharum complex is the accumulation of sucrose in the stems of *Saccharum* spp., albeit at very low levels in the wild *Saccharum* species *S. spontaneum* and *S. robustum* (Table 3.2). Sugarcane species designation has been based on chromosome numbers, floral characters, sugar and fiber content, and stalk diameter (Table 3.2). However, the free intercrossing among the species, the strong influence of environment on the quantitative phenotypic characters, and the wide overlap of measured values do not always allow for clear differentiation. More recently, molecular cytogenetics and genomics have revealed evolutionary relationships among the *Saccharum* species that are more definitive.

Of the four cultivated groups of *Saccharum*, *S. officinarum* L. ($2n = 80$) was the first named and is the primary group for production of sugar. *S. officinarum* accessions have thick stalks with low fiber and high sucrose contents (Table 3.2). *Saccharum barberi* Jeswiet ($2n = 82–124$) in India and *Saccharum sinense* Roxb. ($2n = 88$) in China have been cultivated since prehistoric times, but seldom if ever

Table 3.1 Morphological differences at the generic level between members of the Saccharum complex as described (Mukherjee 1957) and revised (Daniels and Daniels 1975). (modified from (Amalraj et al. 2006))

Character	*Saccharum*	*Erianthus*	*Sclerostachya*	*Narenga*	*Miscanthus*
Root eyes	Two or more rows	Only one row if present	Absent	Absent	One or two rows if present
Bud	Well developed, reproductive	Scaly, except in two species	Absent	Scaly, not reproductive	well developed, reproductive
Dewlap	Present	Absent	Present	Present	Present
Spikelet pair	Sessile and pedicellate	Sessile and pedicellate	Both pedicellate	Sessile and pedicellate	Sessile and pedicellate
Callus hairs	≥2–3 times length of spikelet	~same as spikelet	~0.5 length of spikelet	≤length of spikelet	0.5–2 times length of spikelet

Mukherjee revised the genus *Saccharum* and noted three species (*Erianthus*, *Sclerostachya*, and *Narenga*) that are closely related and interbreeding with sugarcane. He grouped these three species with *Saccharum* referring to the group as the "Saccharum complex" (1954) to indicate a large breeding pool involved in the origin of sugarcane (1957). Daniels et al. added *Miscanthus* to the Saccharum complex postulating it also contributed to the origin of sugarcane (1975)

Table 3.2 Principal characteristics of the six species of *Saccharum* and *Erianthus* and the number of accessions in two of the world collections

Species (chromosome no.)	Common name	Sucrose content (% f. wt.)	Fiber content (% f. wt.)	Stem diameter (cm)	Adaptability	Germplasm Number of accessions US[a]	Germplasm Number of accessions India[b]
S. officinarum ($2n=80$)	Noble	High (18–25)	Low (5–15)	Thick (2.8 ± 0.30)	Tropical	748	764
S. sinense ($2n=110$–120)	Chinese	Medium (12–15)	High (10–15)	Medium (1.4–2.2)	Tropical and subtropical	61	29
S. barberi ($2n=81$–124)	Indian	Medium (13–17)	High (10–15)	Medium (1.7–2.1)	Tropical and subtropical	57	43
S. spontaneum ($2n=40$–128)	Wild	Very low (1–4)	Very high (25–40)	Slender (0.5–0.9)	Tropical through temperate	635	598
S. robustum ($2n=60$–194, usually 80)	Wild	Low (3–7)	Very high (20–35)	Medium (1.1–1.7)	Tropical wetlands	128	145
S. edule ($2n=60$–80)	Edible	Low (3–8)	Low ?	Medium (1.1–1.8)	Tropical	22	16
[c]*Erianthus* ($2n=60$)	Related	Very low	Very high	Slender	Subtropical and temperate	282	

Note. Sucrose, fiber and stem diameter values are either the range or the mean ± SD measured by the Sugarcane Breeding Institute, Coimbatore, India reported in the germplasm catalogs: Sugarcane Genetic Resources I. *Saccharum spontaneum* L. (1983); II. *Saccharum barberi*, Jeswiet; *Saccharum sinense*, Roxb. amend Jeswiet; *Saccharum robustum*, Brandes et Jeswiet ex Grassl; *Saccharum edule*, Hassk. (1985); III. *Saccharum officinarum* L. (1991)

[a]Saccharum germplasm inventory maintained by the USDA, ARS, Miami, Florida; data from, http://www.ars-grin.gov National Plant Germplasm System (GRIN) of the USDA/ARS website March 30, 2010…

[b]Germplasm inventory maintained by the Sugarcane Breeding Institute, Coimbatore, India; data from http://sugarcane-breeding.tn.nic.in/genresources.htm March 30, 2010 and the germplasm catalogs "Sugarcane Genetic Resources vols I, II, III"

[c]*Erianthus* germplasm, classified in GRIN as *Saccharum* spp. (*arundinaceum, bengalense, gigantium*, *brevibarbe*, etc.), are maintained at several germplasm repositories

are used commercially today; they exist primarily in germplasm or garden collections. These two species are sometimes grouped together as a single species, or as historical cultigens, having thin to medium stalks, low to moderate sucrose content, and higher fiber and greater tolerance to stress than does *S. officinarum*. The fourth domesticated species, *S. edule* Hassk. ($2n = 60$ or 80 sometimes up to 122) has an aborted and edible inflorescence and is cultivated from New Guinea to Fiji as a vegetable. Based on its geographical distribution and vegetative morphology, *S. edule* was proposed to be an intergeneric hybrid between either *S. officinarum* or *S. robustum* and a related genus, or to be a mutant of either of these *Saccharum* species (Daniels and Roach 1987; Roach and Daniels 1987). However, molecular data now indicates that *S. edule* may be a series of mutant clones selected from *S. robustum* populations and preserved by humans (Grivet et al. 2006).

Among the wild species, *S. spontaneum* ($2n = 40$–128, with chromosome numbers frequently as multiples of eight) is highly variable with a broad distribution throughout tropical and subtropical regions from Arica to the Middle East, China, Malaysia through the Pacific to New Guinea. *S. spontaneum* accessions exhibit phenotypic, cytological, and cytoplasmic and nuclear DNA sequences that are quite different from those of the other *Saccharum* species. It is a perennial grass, from short bushy types with no stalk, to large-stemmed clones over 5 m in height, but typically with pencil-thin stalks and very low sucrose content (Table 3.2). It is free tillering with robust rhizomes and has contributed towards the development of modern cultivars by conferring resistance to most major diseases, providing vigor and hardiness for increased abiotic stress tolerance (such as cold and drought), increased tillering and improved ratoonability.

The other wild species of *Saccharum*, *S. robustum*, (two cytotypes predominate as $2n = 60$ or 80, but some accessions have chromosome numbers as high as 194) has its center of diversity in New Guinea in the same region as the domesticated *S. officinarum* ($2n = 80$). *S. robustum* has thick stalks and low sucrose content. It is similar to *S. officinarum* and distinguished from *S. spontaneum*, by lack of rhizomes, and thickness and height of stalks including large inflorescences.

Recent genomic data for evaluating genetic diversity within *Saccharum* suggests relationships among accessions that may ultimately produce a definitive classification for the genus. The first molecular evidence came from restriction fragment patterns of nuclear ribosomal DNA that was used to separate accessions of *S. spontaneum*, which showed the widest within-species variation, from accessions of *S. robustum*, *S. officinarum*, *S. barberi*, and *S. sinense* (Glaszmann et al. 1990). RFLP analyses of the mitochondrial genome showed an identical pattern among 18 *S. officinarum* clones and 15 of 17 *S. robustum* clones (D'Hont et al. 1993). RFLP patterns were similar among *S. officinarum*, *S. barberi*, *S. sinense,* and *S. edule*, all of which were distinctively different from *S. spontaneum*. Restriction patterns of the chloroplast genome suggested that, except for *S. spontaneum*, the *Saccharum* species all have the same chloroplast restriction sites (Sobral et al. 1994). RFLP analyses of nuclear genomic DNA confirmed observations about the cytoplasmic genomes that suggested distinctively greater diversity within *S. spontaneum* than within the four other species (Burnquist et al. 1992; Lu et al. 1994a, b; Nair et al. 1999). More recent genomic in situ hybridization analysis supports the hypothesis that *S. barberi*

and *S. sinense* were derived from interspecific hybridization between *S. officinarum* and *S. spontaneum* (D'Hont et al. 2002). These authors conclude that genetic similarities between *S. barberi* and *S. sinense* accessions do not support the taxonomic separation of these two groups into separate species. The species *S. barberi* is not listed in the Kew database GrassBase http://www.kew.org/data/grasses-db.html but is included as *S. sinense*.

1.2 Genome Structure of Modern Cultivars

Modern cultivars are highly polyploid and aneuploid with around 120 chromosomes. They are derived from interspecific hybridization between *S. officinarum* and *S. spontaneum*. As a consequence of the different basic chromosome numbers of *S. officinarum* ($x=10$) and *S. spontaneum* ($x=8$), two distinct chromosome organizations coexist in modern cultivars (see Sect. 3.5 for details on chromosome inheritance in these wide species hybrids). Genomic in situ hybridization (GISH) of total genomic DNA suggests that 10–20 % of modern cultivars chromosomes are inherited in their entirety from *S. spontaneum*; 70–80 % are inherited entirely from *S. officinarum* and around 10 % are the result of recombination between chromosomes from the two ancestral species (D'Hont et al. 1996; Piperidis et al. 2010a; Cuadrado et al. 2004).

Cultivated sugarcane is rare among major crops in being an interspecific aneuploid. The occurrence of chromosomal segment exchanges between *S. officinarum* and *S. spontaneum* is supported by both FISH (D'Hont et al. 1996) and by genetic mapping (Grivet et al. 1996; Hoarau et al. 2001) and disproves the prior assumption (Price 1963, 1965; Berding and Roach 1987) that no recombination occurs between the chromosomes of the two species. Collectively, these data were used to propose a schematic representation (Fig. 3.2) of the genomic organization of modern sugarcane cultivar genomes (D'Hont 2005). Sugarcane's polyploid nature and interspecific origin contribute substantially to high levels of heterozygosity detected among modern cultivars (Lu et al. 1994b; D'Hont et al. 1996; Jannoo et al. 1999a; Lima et al. 2002). On the other hand, the recent origin of modern sugarcane cultivars from a small germplasm base, results in strong linkage disequilibrium with some haplotypes conserved in segments extending for at least 10 cM (Jannoo et al. 1999b), far greater than in most other crops.

2 Secondary and Tertiary Gene Pools, Germplasm Resources

2.1 Related Genera. Saccharum, Erianthus Sect.

Ripidium, *Sclerostachya* ($2n=30$), *Narenga* ($2n=30$), and *Miscanthus* ($2n=38$ and 40) were assembled into the "*Saccharum* complex" as a closely related interbreeding group (Mukherjee 1957; Daniels et al. 1975). *Erianthus* sect. *Ripidium* includes

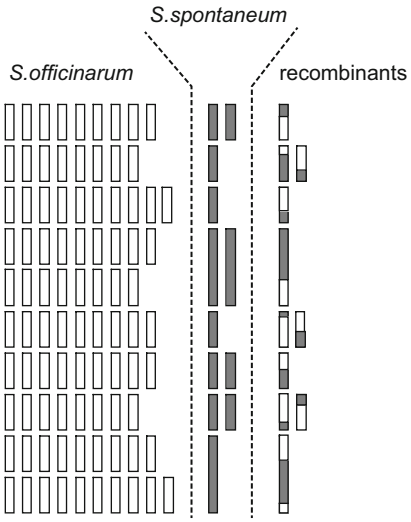

Fig. 3.2 Schematic of the genome of a typical modern sugarcane cultivar. Each *bar* represents a chromosome; *open boxes* represent regions originating from *S. officinarum* and *shaded boxes* regions from *S. spontaneum*. Chromosomes aligned in the *same row* are hom(oe)ologous and represent a homology group (HG). Chromosomes assembled *vertically* correspond to monoploid genomes (MG) of *S. officinarum and S. officinarum*. The key characteristics of this genome are the high level of ploidy, the aneuploidy, the bispecific origin of the chromosomes, the existence of structural differences between chromosomes of the two origins, and the presence of interspecific chromosome recombinants (From Fig. 2 (D'Hont et al. 2008). Used with kind permission of Springer Science and Business Media)

chromosome numbers of $2n = 20$, 30, 40, and 60 with a basic chromosome set of $x = 10$, the same basic number as favored for *Saccharum* (Whalen 1991). This group of related genera have a number of traits desired by sugarcane breeders for improving cultivar performance including a wide environmental distribution due to tolerance to environmental stresses of cold and drought, vigor, good ratooning, and disease resistance (Berding and Roach 1987).

Sugarcane breeders have long tried to introgress agronomic characteristics of *Miscanthus* and *Erianthus arundinaceus* into sugarcane hybrids. However, it has been difficult to produce fertile progeny and to identify true hybrids involving *E. arundinaceus* on the basis of morphological characters (Grassl 1965). Molecular diagnostic tools including species-specific DNA markers and GISH have been used in attempts to identify *Saccharum* × *Erianthus* hybrids at the seedling stage and to follow the introgressed genes into later generations (D'Hont et al. 1995; Alix et al. 1998, 1999; Piperidis et al. 2000). GISH allowed analysis of the chromosome complement of intergeneric hybrids involving *Erianthus* and *Miscanthus* (D'Hont et al. 1995) and revealed that chromosome elimination occurs in *Saccharum* × *E. arundinaceus* hybrids (D'Hont et al. 1995; Piperidis et al. 2000).

GISH revealed a high contrast between the chromosomes of the two genera in *Saccharum* × *E. arundinaceus* hybrids as compared to that of *S. officinarum* × *S. spontaneum* hybrids. Since GISH is based on the presence of species-specific

repeated sequences that evolved quickly during speciation, the greater contrast between *Saccharum* and *Erianthus* could reflect a greater genetic distance between these two genera than might be expected based on morphological characters. This difference in chromosome structure may explain the occurrence of chromosome elimination and the difficulties encountered by breeders attempting to exploit this genus. Several *Erianthus* and *Miscanthus* specific repeated sequences were cloned. Their distribution on the chromosomes was analyzed by FISH and revealed two subtelomeric families (Alix et al. 1998), one centromeric family, and one family apparently dispersed along the genome (Alix et al. 1999).

More recently AFLP markers clearly identified hybrids between *S. officinarum* or sugarcane cultivars and *Erianthus rockii* (Aitken et al 2006). Both crosses produced progeny all showing $n+n$ inheritance. All of the progeny from the *S. officinarum* cross were hybrids but only 10 % of the progeny from the sugarcane cultivars were hybrids of *E. rockii,* the remaining 90 % were identified as selfs.

2.2 Germplasm Resources

Sugarcane breeders have long realized that a large diverse germplasm collection is essential for sustained crop improvement. At least 31 separate collecting expeditions across the complete natural distribution of *Saccharum* species were made from 1892 through 1985, to collect genotypes focusing on those that were resistant to pests and diseases, were highly productive, or had high sugar content (Berding and Roach 1987). Clones from these collections have been deposited in the two world collections, one maintained in India and one in the USA. These collections serve as genetic reservoirs to be used in breeding new cultivars for specific agronomic needs and to broaden the genetic base of commercially grown cultivars. The reported number of accessions for each species in the World Collection of Sugarcane and Related Grasses located at the India and US sites are listed in Table 3.2.

2.3 Passport and Descriptor Information

Passport data, including taxonomic designation and information about where an accession was collected and phenotypic descriptor data are available from the Germplasm Resources Information Network (GRIN) database maintained by the National Plant Germplasm System (NPGS) of the USDA at http://www.ars-grin.gov/cgi-bin/npgs/html/crop.pl?101. The 102 descriptors are classified into eight categories, with the largest categories being morphology (69 descriptors) and disease reactions (19). These data are useful for classification of accessions, but they are subject to environmental influences and most have a significant genotype × environment interaction.

2.4 Phenotypic Evaluation

The evaluation of germplasm held in collections is a high priority. This information facilitates the use of germplasm and makes more efficient management possible. Evaluation is a lengthy and costly process of examining accessions for traits of significance; however, it adds tremendous value to germplasm collections.

Clones of *S. sinense*, S. *barberi*, and *S. robustum* were evaluated for agronomic and quality characters and to estimate the genetic diversity within and between the populations. Thirty clones of each species were evaluated in two environments for eight phenotypic characteristics. Significant differences were found between the three species as well as for clones within species. The genetic repeatability for every character, except stalk number, was high, indicating that this information would be useful for breeders interested in using the material in commercial crosses (Brown et al. 2002). Additional phenotypic evaluation continues to characterize this germplasm for quality-related characteristics (sucrose, starch, etc.) and to estimate its tolerance to environmental stresses (freezing temperatures, diseases, etc.).

2.5 Core Collections

Potential usefulness for establishing core collections of *Saccharum* species has been analyzed for both the India and US world collections. With the exceptions of *S. spontaneum* and *S. officinarum*, the numbers of accessions of the other species that have been characterized are so few that Tai and Miller (2001) considered the entire US collection of those limited species to function as cores. Workers in India analyzed their collections of both *S. spontaneum* and *S. officinarum*, while workers in the USA limited their analysis to their collection of *S. spontaneum*. In the USA, Tai and Miller (2001) evaluated 11 methods using 11 quantitative traits to estimate the number of randomly selected accessions needed for a representative core to be approximately 75. Although the authors did not suggest any one core as the best, they did list members of the core based on cluster analysis within each geographic region based on retained principal components for morphological traits and random selection of entries within each cluster. Workers in India analyzed a set of 21 qualitative and 10 quantitative descriptors on 617 accessions of *S. spontaneum* and found most characters would be represented in a core size of about 60 randomly sampled accessions (Amalraj et al. 2006). In a separate study workers in India analyzed a set of 27 qualitative morphological descriptors for computing the Shannon–Weaver diversity index and a list of 10 quantitative descriptors for principal component analysis of 690 accessions of *S. officinarum* to find a core optimum of about 164 accessions (Balakrishnan et al. 2000). Reports from both world collections emphasized the smaller diversity in *S. officinarum* than in *S. spontaneum*. Although the potential for establishing core collections of *Saccharum* has been shown in both world collections, both collections continue to be preserved in their entirety.

3 Evolution and Improvement of Sugarcanes

3.1 The Origin of Sugarcane

Sugarcane prehistory apparently occurred across a vast area of Southeast Asia including the Malayan Archipelago, New Guinea, India, and some of the island groups of Melanesia. The preponderance of evidence indicates that domestication of sugarcane probably occurred in New Guinea with the selection of *S. officinarum* from the wild species *S. robustum* (Brandes 1956; Daniels and Roach 1987; Grivet et al. 2006). Hypotheses about possible contributions to sugarcane of genera other than *Saccharum*, specifically *Erianthus*, *Sclerostachya*, *Narenga*, and *Miscanthus* (Barber 1920; Jeswiet 1930; Parthasarathy 1948; Brandes 1956; Mukherjee 1957) were based on cytology and breeding evidence, morphology, and overlapping geographical distribution. These hypotheses were reviewed in Daniels and Roach (1987) who produced the synopsis supporting the scenario developed by Brandes (1956) which has been supplemented by molecular data (Grivet et al. 2006) and presented as Fig. 3.4.

Cultivated sugarcanes of today are complex interspecific hybrids primarily between *Saccharum officinarum*, known as the noble cane, and *Saccharum spontaneum* with contributions from *S. robustum*, *S. sinense*, and *S. barberi*. Early authorities hypothesized additional contributions from related grasses of the *Saccharum* complex (Brandes 1956; Daniels and Roach 1987; Roach and Daniels 1987; Sreenivasan et al. 1987) but such proposals are not supported by the limited molecular evidence available (Irvine 1999, Grivet et al. 2006). Based on vegetative characters and native distribution, the species *S. officinarum*, with high sucrose content, is believed to have been derived from *S. robustum* in New Guinea (Brandes 1956, 1958). It has been postulated (Brandes 1956) and widely accepted from various evidence (Daniels and Roach 1987; Roach and Daniels 1987) that *S. officinarum* spread during prehistoric times from New Guinea to Indonesia, Malay, China, India, Micronesia, Polynesia, and by A.D. 500 from Indonesia to southern Arabia and East Africa. As detailed below, *S. barberi* and *S. sinense* were likely derived from interspecific hybridization between *S. officinarum* and *S. spontaneum* (Grivet et al. 2006) and possible introgression from other species (Brown et al. 2007). The distribution of *S. officinarum* from Polynesia to Hawaii probably took place with native migrations around A.D. 500–1000.

3.2 Origin of S. barberi and S. sinense

Sugarcanes indigenous to North India and China and cultivated from prehistoric times are referred to as *S. barberi* ($2n = 81–124$) and *S. sinense* ($2n = 110–120$), respectively. *S. sinense* cultivated in China and Pansahi India was used for chewing

as well as for sugar production, whereas the thinner, harder stalks of *S. barberi* cultivated in northern India may have been primarily for crushing. The two cultivated sugarcanes were probably the result of natural hybrids of *S. officinarum* and *S. spontaneum* with other genera about 1000 BCE. *S. barberi* subsequently spread from India to the Middle East, Mediterranean, and to the New World beginning with the second voyage of Columbus in 1493. Today, *S. sinense* and *S. barberi* exist only in collections (Stevenson 1965; Blume 1985; Rossi et al. 1987; Heinz et al. 1994).

Leading hypotheses on the origins of *S. barberi* and *S. sinense* (reviewed by Daniels and Daniels 1975; Paton et al. 1978; Daniels and Roach 1987; Roach and Daniels 1987; D'Hont et al. 2002, Grivet et al. 2006) are that: (1) *S. barberi* and *S. sinense* arose from hybridization of *S. officinarum* with *S. spontaneum* in India and China; (2) *S. barberi* was developed from *S. sinense* in India; and (3) *S. barberi* and *S. sinense* arose through introgression between *S. officinarum*, *S. spontaneum*, or other genera such as *Erianthus* and *Miscanthus* (Brown et al. 2007). Whalen (1991) and Artschwager and Brandes (1958) considered *S. barberi* to be a horticultural variant of *S. sinense*, as does the Kew database GrassBase http://www.kew.org/data/grasses-db.html even though it lists 37 species in the genus *Saccharum*.

The hypotheses about the origin of *S. sinense* and *S. barberi* were tested by genomic in situ hybridization (GISH) performed using *S. spontaneum* total genomic DNA and *S. officinarum* total genomic DNA as probes on chromosome preparations of genotypes representative of *S. barberi* and *S. sinense*. In all clones analyzed, GISH clearly identified two distinct populations of chromosomes or chromosome fragments, thus revealing the interspecific origin of *S. barberi* and *S. sinense* (D'Hont et al. 2002). GISH showed no genomic regions lacking color, nor was there a third color pattern that would have been the case if a third species were involved, especially if it belonged to another genus. For example, GISH performed on intergeneric hybrids between *S. officinarum* × *Erianthus* or *S. officinarum* × *Miscanthus* showed that total genomic DNA of one genus gave a very weak hybridization signal on the other genus (D'Hont et al. 1995; Piperidis et al. 2000 and unpublished results of these workers). These results are corroborated by the absence of *Erianthus* or *Miscanthus* genus specific sequences in *S. barberi* and *S. sinense* on Southern hybridization patterns (Alix et al. 1998, 1999). These results, together with cytoplasmic (D'Hont et al. 1993) and nuclear molecular marker analyses (Glaszmann et al. 1990; Lu et al. 1994a), are in agreement with the origin of both *S. barberi* and *S. sinense* from hybridizations between *S. officinarum* (female) and *S. spontaneum* (male).

The proportion of chromosomes from the two species was variable in the clones studied with 61 % *S. officinarum*: 39 % *S. spontaneum* for $2n = 82$ clones, 68 %: 32 % for a clone with $2n = 91$, and 66 %: 33 % for a clone with $2n = 116$. From 0 to 4 chromosomes per cell appeared to result from interspecific intrachromosomal exchanges (D'Hont et al. 2002). Considering the frequency of such exchanges in modern cultivars, this indicates that a very small number of meiotic events must have occurred since interspecific hybridization. Further RFLP analyses indicated that the *S. barberi* and *S. sinense* clones are clustered into a few groups, each derived from a single interspecific hybrid that has subsequently undergone somatic mutations.

These groups correspond quite well with those already defined based on morphological characters and chromosome numbers (reviewed by Daniels et al. 1991). However, the calculated genetic similarities do not support the existence of two distinct taxa. The "North Indian" and "Chinese" sugarcanes thus represent a set of horticultural groups rather than established species (D'Hont et al. 2002).

3.3 Cultivated Noble Canes for Sugar Production

As mentioned above, sugarcane culture spread from India to the Middle East, Mediterranean, and to the New World in 1493, well before establishment of classification by Linnaeus in 1753 and before the discovery of "noble" canes in the islands of the South Seas. The first of the noble canes left Tahiti with Bougainville in 1768, eventually arriving in the Caribbean in 1789 (Deerr 1921, 1949; Machado et al. 1987) The sugarcane that spread across the Mediterranean to the New World was the Indian cultivar known as "Creole" in French, "Criola" in Spanish, or "Crioula" in Portuguese. "Creole" was quickly replaced in cultivation by the noble cultivar "Otaheite" when it was brought to Jamaica from Tahiti by Bligh in 1793 (Rossi et al. 1987). From there it was distributed throughout the Caribbean and the Americas. Original noble canes collected from the Pacific Islands quickly replaced the less productive Indian varieties and were the only source of cultivars for plantations for the world's sugar production for over a hundred years. Before sugarcane breeding programs were started in 1888, the most important noble cultivars were the "Otaheite" (Bourbon, Lahaina) of Tahiti, "Cheribon" (Louisiana Purple) of Java, and "Caledonia" of New Hebrides. "Bourbon" was extremely susceptible to root rot, mosaic, and gumming disease; "Cheribon" to sereh, mosaic, and root rot; and "Caledonia" to mosaic (Edgerton 1958; Stevenson 1965). These initial cultivars were replaced by new hybrids selected from emerging sugarcane breeding programs (Fig. 3.3). Today, clones of *S. officinarum* are in breeding collections and/or cultivated as garden canes for chewing.

The first sugarcane breeding programs began in Java and Barbados in 1888, following the observations independently in Java (1858) and Barbados (1859) that sugarcane was capable of producing viable seed (Stevenson 1965; Kennedy and Rao 2000). Varieties produced by the Proefstation Oost Java, identified as (POJ) varieties, became foundational for germplasm development in other countries which soon established their own breeding stations to produce locally adapted varieties. Notable among the early sugarcane breeding efforts for producing varieties with wide adaptation was Coimbatore, India (1912) that developed Co and NCo varieties (Fig. 3.3). In the sugarcane breeding history that follows we describe these changes in four stages: (1) breeding among noble canes (*S. officinarum*) to produce noble cultivars, (2) breeding through nobilization, i.e., interspecific hybridization with backcrossing to noble cultivars to produce nobilized cultivars, (3) breeding of nobilized canes to produce hybrid cultivars, and (4) breeding to broaden the genetic base.

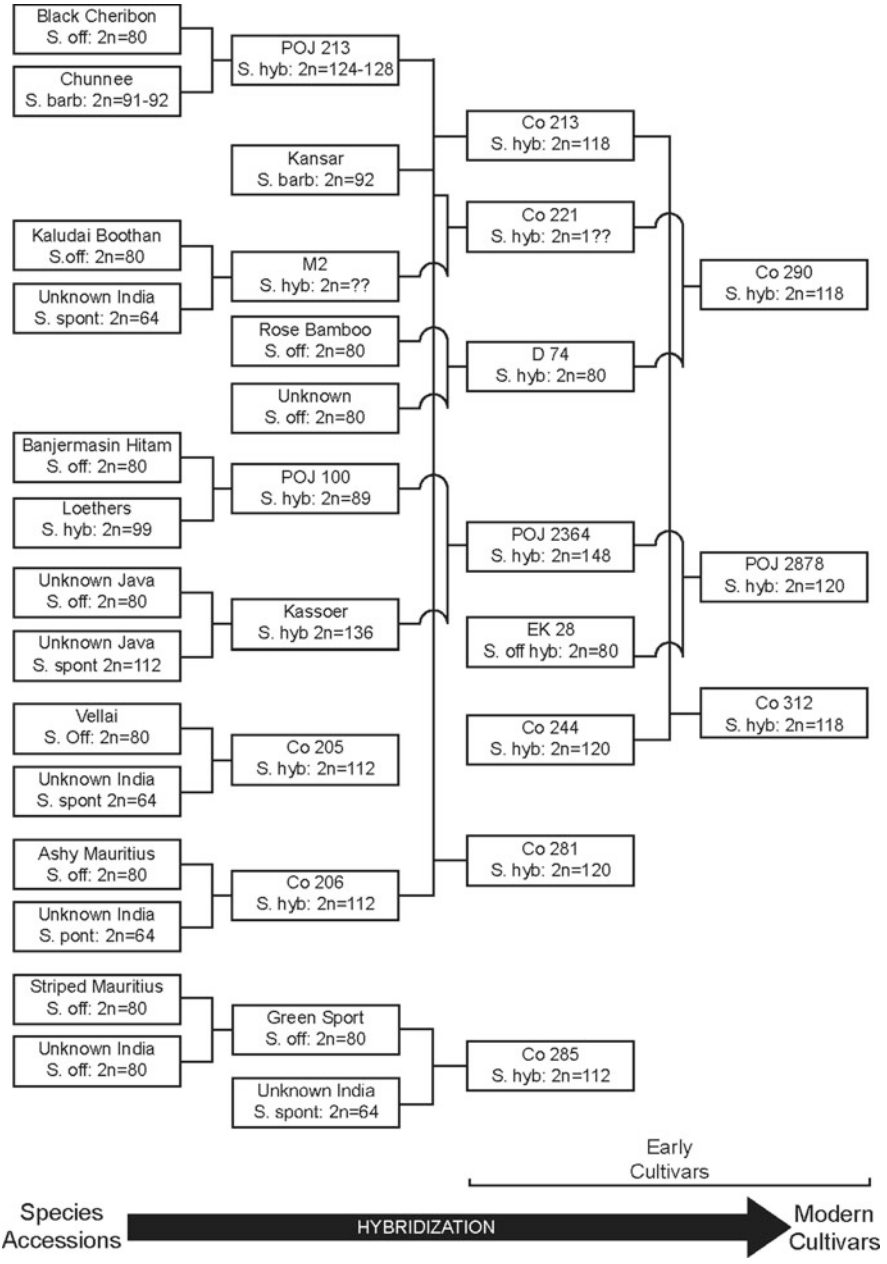

Fig. 3.3 Sugarcane Hybrid Foundations. Accessions of *Saccharum* species and the early crosses among them that serve as the foundation for all modern cultivars of sugarcane. The early selections from crosses by Proefstation Oost Java are named as a series of POJ varieties and those conducted in Coimbatore, India are named as a series of Co varieties. The accessions of *S. officinarum* ($2n=80$), *S. spontaneum* ($2n=64$–112) and *S. Barberi* ($2n=92$) were named where discovered and served as the original sugarcane cultivars. Early generation crosses among *S. officinarum* produced a new series of *S. officinarum* hybrids (S. off hyb: $2n=80$), or when a different species was crossed to *S. officinarum* the lines produced are *Saccharum* hybrids (S. hyb: $2n=112$–148)

3.4 Breeding of Noble Canes to Produce Noble cultivars

Progenies of open-pollinated noble canes were selected for sugar production. Each selected seedling was assigned a call sign followed with a seedling number. "Otaheiti" ("Lahaina," "Bourbon") produced the EK seedlings in Java, "H109" in Hawaii, and "B716" in Barbados. In Queensland, "Q813" came from Badila, the famous chewing cane, originating in New Guinea. These selected noble cultivars were important for sugar production in the early 1900s. The original noble canes and selected noble progenies were found susceptible to disease and insects and limited to particular tropical environments. Breeders soon realized that the genetic base of the noble canes needed to be broadened to improve their adaptabilities and disease and insect resistance (Stevenson 1965).

3.5 Breeding Through Nobilization to Produce Nobilized Cultivars

Nobilization is the pollination of noble cane *S. officinarum* with its wild relative *S. spontaneum* followed by repeated backcrosses to noble canes. The wild relative were considered "nobilized" through the breeding process and the selected hybrid progenies are referred to as "nobilized canes" (Bremer 1961). A key event of early nobilization breeding was the production of the nobilized cultivar, "POJ2878" in 1921 (Fig. 3.3) that became the most universal breeding cane ever developed and is found in the pedigrees of almost all of the dominant cane varieties grown around the world.

The first step of the nobilization process involved "doubling" of the *S. officinarum* gametic chromosome number to the somatic chromosome number in the fertilized egg with the addition of the gametic chromosome number of the wild clones of *S. spontaneum* used as males. The mechanism to explain maternal transmission of diploid chromosome numbers seems to involve the fusion of daughter nuclei after the second meiotic division in the innermost megaspore dyad cell of *S. officinarum* (Narayanaswami 1940). However, doubling also occurs in crosses involving the species *S. sinense* as shown by Price (1957) and even modern cultivars (Piperidis et al. 2010b). Subsequent steps in nobilization involved backcrossing the F1 to another noble cane, where there could be a second doubling of maternal chromosomes, and then crossing the F2 once again to a noble cane, at which time normal $n+n$ transmission seems to be the norm.

More than 90 % of the accessions classified as *S. officinarum* have $2n=80$, $x=10$ chromosomes whereas the most frequent chromosome number in *S. spontaneum* is $2n=64$, $x=8$ (Panje and Babu 1960; Irvine 1999). Using these two chromosome

complements as an example, one can envision nobilization in the following simplified crossing scheme where NN = noble, $2n = 80$; SS = spontaneum, $2n = 64$.

Female ×	Male →	Progeny (chromosomes)	(% Spontaneum)
NN(80) ×	SS(64) →	F_1 (80 + 32 = 112)	(S % = 29 = 100 × 32/112)
NN(80) ×	F1(112) →	BC_1 (80 + 56 = 136)	(S % = 12)
NN(80) ×	B1(136) →	BC_2 (40 + 68 = 108)	(S % = 7)

Progeny of F_1 and BC_1 have the nonreduced somatic complement ($2n$) of the female parents plus the gametic number (n) of the male. Most cultivated nobilized canes (BC_2s, BC_3s, etc.) have 100–130 chromosomes with about 5–10 % from *S. spontaneum* (Fig. 3.4). Clones with chromosome numbers outside of this range are rarely suited for commercial production

Following efforts to achieve interspecific crosses between *Saccharum officinarum* and *S. spontaneum*, early sugarcane breeders realized that resultant F_1 progeny were distinctively more robust than either parent. When *S. officinarum* clones were used as the female parent, progeny tended to be larger stalked, higher in sucrose levels, and generally more vigorous than when *S. spontaneum* clones were used as the female parent. Reciprocal differences in vigor were eventually explained by the cytological phenomenon of a high frequency of "$2n + n$" progeny in *S. officinarum* (female) × *S. spontaneum* (male) crosses (Bremer 1923).

In Coimbatore, India, nobilization of *S. barberi* and *S. spontaneum* with *S. officinarum* produced the famous early nobilized tri-species hybrids of "Co" seedlings (Fig. 3.3) that gained wide acceptance in subtropical regions in India, South African, Australia, Louisiana, Argentina, and Brazil. The "Co" cultivars also were used on the poorer soils and under marginal growth conditions in the tropics. After 1925–1930, nobilization breeding was seldom used (Stevenson 1965; Simmonds 1976; Ethirajan 1987).

3.6 *Breeding of Nobilized Canes to Produce Hybrid Cultivars*

Crosses among nobilized canes in the 1930s produced many important hybrid cultivars for sugar production in the next three decades (Fig. 3.3). Breeding of "POJ2878" with other nobilized POJ canes produced cultivars "POJ3016" and "POJ3067." Together they occupied more that 85 % of the cane area of Java in 1960. Crossing of "Co312" and "POJ2878" produced Hawaii's most important cultivar, "H32-8560," which was responsible for 65 % of the cane area of Hawaii in 1945. "POJ2878" × "Co290" produced "Co419" for the tropical area of India. The cross of "Co421" × "Co312" was made in Coimbatore in 1938 to produce progeny of the cross that was grown in Natal, South Africa, in the same year. One of the progeny selected in 1939, "NCo310," became the most important cultivar of the world in the 1950s and 1960s (Anonymous 1945; Nuss and Brett 1995). Even as late as the 1980s, "NCo310" still ranked first in growing area in Japan, Texas (USA), and Uruguay; second in Malawi and Gabon; third in Mexico, and fourth in Ecuador (Tew 1987).

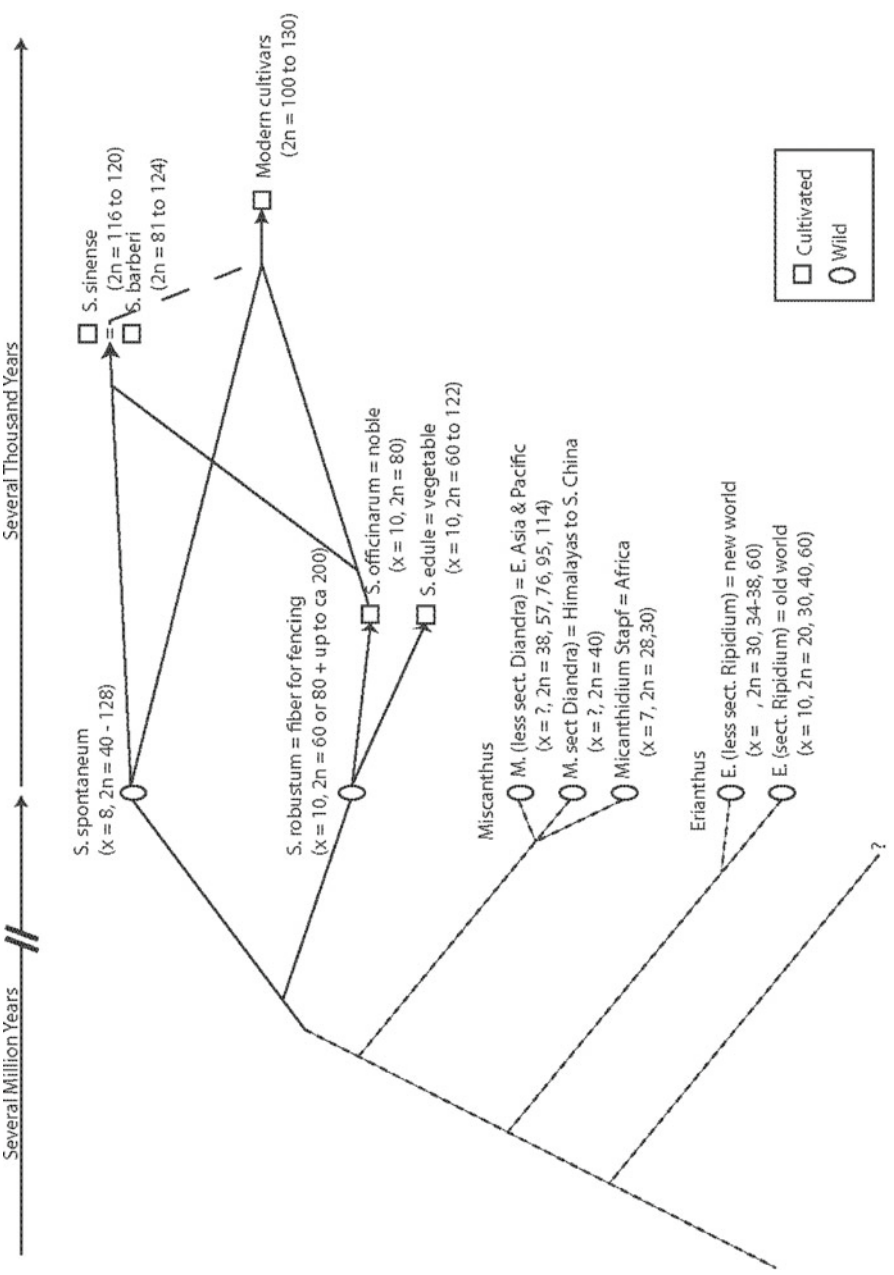
Several Million Years
Several Thousand Years
S. spontaneum
(x = 8, 2n = 40 - 128)
S. sinense
(2n = 116 to 120)
S. barberi
(2n = 81 to 124)
Modern cultivars
(2n = 100 to 130)
S. robustum = fiber for fencing
(x = 10, 2n = 60 or 80 + up to ca 200)
S. officinarum = noble
(x = 10, 2n = 80)
S. edule = vegetable
(x = 10, 2n = 60 to 122)
Miscanthus
M. (less sect. Diandra) = E. Asia & Pacific
(x = ?, 2n = 38, 57, 76, 95, 114)
M. sect Diandra) = Himalayas to S. China
(x = ?, 2n = 40)
Micanthidium Stapf = Africa
(x = 7, 2n = 28,30)
Erianthus
E. (less sect. Ripidium) = new world
(x = , 2n = 30, 34-38, 60)
E. (sect. Ripidium) = old world
(x = 10, 2n = 20, 30, 40, 60)
?
Cultivated
Wild

Fig. 3.4 Hypothetical pathway for sugarcane evolution and domestication based on molecular data indicating only members of the *Saccharum* clade contributed directly to sugarcane cultivars through allopatric speciation of *S. spontaneum* west of Sulawesi and *S. robustum* east of Sulawesi (Grivet et al. 2006). *Miscanthus* and *Erianthus*, previously proposed as contributing to sugarcane are recognized as sharing common ancestors. Following *Saccharum* speciation, *S. robustum* in New Guinea contributed to cultivars of *S. officinarum* for sugar, *S. edule* for vegetables, and *S. robustum* for fencing and construction. *S. officinarum* was moved by humans from the tropics to more temperate India and China where it hybridized to native *S. spontaneum* giving rise to *S. barbari* and *S. sinense* as locally adapted sugarcane cultigens. Chromosome numbers from compilation by Daniels and Roach (1987). Strongest evidence for the evolution and domestication are indicated with *solid lines*; weaker evidence is indicated with *dashed lines* (Modified after Grivet et al. 2006 and D'Hont et al. 2008)

Commercial cultivars and hybrids from advanced stages of selection have been the main breeding materials for the development of current cultivars since the 1950s. The names of the current cultivars, rank in percent of area occupied, immediate parents, and breeding stations of the world are listed in the World Sugarcane Variety Census—Year 2000 (Tew 2003).

3.7 Breeding to Broaden the Genetic Base

Modern cultivars are essentially derivatives of no more than 15–20 nobilized cultivars that in turn trace back to the initial nobilized genetic base developed in Java and India (Roach 1989). Genetic diversity in today's advanced breeding populations is expected to be somewhat narrower than that in the initial germplasm following more than 100 years of directional selection (Walker 1987). Attempts to increase this narrow base, called the base broadening program (BB-program) were started in Barbados in 1965 using clones different from those initially used in Java and India. The BB-program started with nobilization crosses followed with hybridization of nobilized canes. The BB-program has produced many semicommercial type clones that are being incorporated into the gene pool of advanced breeding populations since the late 1980s (Kennedy and Rao 2000). Other countries have tried BB-programs over the past 50 years by crossing wild canes with their commercial cultivars. However, none of these efforts was as long term nor as broad based as the BB-program of Barbados and Louisiana

Our inability to trace or follow the incorporated germplasm into the germplasm of the advanced breeding population through visual selection is perhaps the main reason for failure of base broadening programs. Large favorable genetic variation exists among clones of *Saccharum* species (Tai and Miller 2002). What is missing is a breeding tool to assist breeders in incorporating useful genes from any source into the gene pool of the advanced cultivars. Recent developments in biotechnology are beginning to yield information and technologies that undoubtedly will help the breeders to broaden the gene pool of their advanced breeding populations and produce higher yielding cultivars in the future.

4 Evolution and Improvement of Energy Canes

4.1 Basis for Biomass Breeding

Sugarcane has long been recognized as one of the world's most efficient crops in converting solar energy into chemical energy harvestable as biomass. The theoretical upper limit for sugarcane biomass (total solids) production under a 365 day growing season in the tropics has been estimated to be 177 t/ha-year (Waclawovsky et al. 2010). The maximum sugarcane biomass yields that have been achieved in several

countries are just over half of the calculated theoretical maximum. Though limited to tropical and subtropical climates, sugarcane is a worldwide crop, commercially grown in 2004 in more than 80 nations. On a fresh weight basis, a larger mass of this crop is harvested and transported for processing than any other crop in the world, exceeding 1.6 billion tonnes (1,600 Mt) in 2008; sugarcane production in Brazil alone reached 649 Mt (FAOSTAT Database: (http://apps.fao.org), from which 32.4 Mt of sugar and 24.5 billion liters of ethanol were produced (http://en.wikipedia.org/wiki/Ethanol_fuel_in_Brazil)).

There are very few agronomic crops that rival sugarcane in energy conversion efficiency. In comparison to sorghum and maize, sugarcane requires less nitrogen input to achieve a full crop and it is generally cultivated as perennial crop, both considerations contributing to its exceptional energy efficiency. The "energy ratio" or "net energy balance" of sugarcane has been frequently cited as being far superior to most agronomic crops—in the order of 3–8 units energy output for each unit of energy input (Goldemberg 2008; Yuan et al. 2008).

4.2 Biomass Breeding Background

Sugarcane breeding and selection has traditionally focused on traits that result in consistently high total sugar yield and sugar content at harvest and in minimal planting, crop maintenance, and harvesting costs. Milling considerations also come increasingly into play as selections reach the more advanced stages of evaluation. Mills require that dry fiber levels be within a fairly narrow range, usually 9–12 % of fresh cane weight. The mills need sufficient fiber to efficiently process cane and to have enough bagasse as a fuel source for their boilers to meet their own electrical needs, including whatever electrical generation commitments they may have to the surrounding community. High fiber levels result in reduced milling efficiency. Penalties and incentives from the mill have tended to reinforce a stringent fiber range that breeders have traditionally worked within toward genetically improving sugarcane for commercial sugar production.

Alexander (1985), a pioneer energy cane advocate, argued that biomass yields could be two- to threefold that of present expectations with a reorientation of cane management involving (1) utilization of aggressive high-fiber genotypes, (2) utilization of the whole plant including tops and leaves that have traditionally been burned just prior to harvest, and (3) growth orientation from planting to harvest.

Sugarcane base broading efforts (see Sect. 3.7), primarily involving *S. spontaneum*, have resulted in germplasm that is well suited to most biomass breeding strategies. Early-generation germplasm is generally more vigorous, better ratooning, adapted to more temperate environments, and more genetically diverse than the commercial sugarcane germplasm from which they were derived. While early-generation hybrids generally do not meet sugar and fiber levels required in commercial sugarcane, such standards are no longer prohibitive where total energy capture is sought.

4.3 Biomass Breeding Strategies

Tew and Cobill (2008) described three distinctive strategies that have been followed in the development of so-called energy canes.

Strategy 1: Status quo. Brazil, far and away the largest producer of ethanol from sugarcane in the world, continues to focus on breeding higher yielding sugarcane for ethanol production with essentially the same quality characteristics as those found in traditional sugarcane. The Brazilian industry is best able to shift from an ethanol emphasis toward refined sugar emphasis and vice versa, depending on economics, following this strategy. Brazil's sugarcane yields on a per-hectare basis have increased substantially since 1986 as a result of genetic and technological improvements, attesting to the success of this strategy (Xavier 2007).

Strategy 2: Type I energy cane—focus on sugar and fiber. The objective in this strategy is to identify clones that possess the same or nearly the same sugar quality characteristics found in traditional sugarcane, while possessing elevated fiber content. The energy cane concept advocated by Alexander (1985) and the varietal model, US 67-22-2, that he used in Puerto Rico best conforms to this strategy. The goal is to maximize total energy capture efficiency, but within limitations that allow sugarcane to continue to be used as a sugar crop. In recent years the USDA Sugarcane Research Facility located at Houma, Louisiana has released some high-fiber *Saccharum* cultivars that fit into the Type I energy cane model.

Strategy 3: Type II energy cane—focus on fiber only. The objective of this strategy is to breed, select, and cultivate cane primarily or solely for its fiber content. Type II energy cane would be used as a feedstock source for generation of electricity and for production of cellulosic biofuel. Type II energy cane cultivars would likely have greatest appeal in mild temperate environments beyond the range that traditional sugarcane can be successfully grown in, where freezing of above-ground tissue is not only expected, but may even be desired in order to facilitate desiccation prior to harvest. Obviously, winter survival would be a critically important trait in this scenario. The Louisiana State University AgCenter and the USDA have both released high-fiber *Saccharum* cultivars that fit into the Type I energy cane model for use in US biofuels industry in southeastern US. The range of their winter survival outside the Louisiana sugar belt has yet to be determined.

5 Other Uses

Sugarcane has several traits that contribute to its potential to become a key crop for transition from petrochemical-based to bio-based economies. Sugarcane is fast-growing, it produces a large biomass, partitions carbon into sucrose at up to 42 % of the dry weight of the stalk, and has a mobile pool of hexose sugars through most of its life cycle. The paucity of viable seed from commercial cultivars greatly reduces the potential

for "transgene escape" from an agricultural crop engineered to produce industrial chemicals. In addition, sugarcane regrows from the underground portions, or stools, left in the ground after harvesting and therefore can be harvested multiple times, reducing the potential for soil erosion and the costs of planting for crop establishment. Sugarcane has relatively few pests and diseases and as a C4 plant it is an efficient user of water and nitrogen.

Several groups have been involved in the development of sugarcane as a biofactory for the transgenic production of high value proteins, plastics, and carbohydrates. A group in Hawaii (Wang et al. 2005) focused on using transgenic sugarcane to produce human granulocyte macrophage colony stimulating factor (GM-CSF), which is used in clinical applications for the treatment of neutropenia and aplastic anemia. This work culminated in the first US field trial of transgenic sugarcane to produce a human pharmaceutical protein. Accumulation of the GM-CSF protein was up to 0.02 % of total soluble protein. Research at Texas A&M University used sugarcane to produce pharmaceutical-grade human structural proteins for human therapeutics, likewise accumulating very low levels of the desired product (Holland-Moritz 2003).

Groups in Australia tested the ability of sugarcane to be a biofactory for the production of high value products other than proteins. Brumbley and coworkers (Brumbley et al. 2002) engineered the genetic pathway for poly-3-hydroxybutyrate (PHB) (Schubert et al. 1988; Peoples and Sinskey 1989a, b) into sugarcane. Because PHAs have thermoplastic properties and are biodegradable, they are attractive alternatives to petrochemically derived plastics. Brumbley and coworkers (Brumbley et al. 2002) targeted the products from the *Ralstonia eutropha* PHB biosynthetic pathway to several subcellular compartments of sugarcane. In the best producing line, polymer accumulated in the leaves at 1.2 % of dry weight, 0.4 % in the stem rind, and 0.004 % in the stem pith. Analysis of height, weight, and sugar levels revealed no significant difference between transgenic and wild type lines.

Sugarcane also was evaluated as a production platform for a major component of liquid crystal polymers (LCP), *p*-hydroxybenzoic acid (pHBA), using two different bacterial proteins, a chloroplast-targeted version of *Escherichia coli* chorismate pyruvate-lyase (CPL) (Siebert et al. 1996) and a *Pseudomonas fluorescens* 4-hydroxycinnamoyl-CoA hydratase/lyase (HCHL) (Gasson et al. 1998). Both approaches provide a one-enzyme pathway from a naturally occurring plant intermediate. McQualter and coworkers (McQualter et al. 2005) demonstrated that HCHL is the superior catalyst for production of pHBA in sugarcane leaf and stem tissue. *p*-Hydroxybenzoic acid was quantitatively converted to glucose conjugates by endogenous UDP-glucosyltransferases and was presumably stored in the vacuole. The largest amounts detected in sugarcane leaf and stem tissue were 7.3 and 1.5 % dry weight, respectively, yet there were no discernible phenotypic abnormalities in any of the sugarcane lines tested (McQualter et al. 2005). However, as a result of diverting carbon away from the phenylpropanoid pathway, there was a reduction in leaf chlorogenic acid, subtle changes in lignin composition, and an apparent compensatory upregulation of phenylalanine ammonia-lyase (McQualter et al. 2005). In addition to pHBA, vanillic acid also was produced in sugarcane plants expressing the HCHL genes (McQualter et al. 2005).

Other work in Australia focused on developing sugarcane as a platform for the production of higher value isomers of sucrose, the naturally stored sugar of sugarcane. Birch and coworkers engineered an efficient sucrose isomerase from the bacterium *Pantoea dispersa* (Wu and Birch 2007) with a monocot promoter and a vacuolar targeting sequence (Gnanasambandam and Birch 2004) and used this construct to transform sugarcane for the production of isomaltulose. Isomaltulose accumulated in sugarcane stem storage tissues of transformed plants without any decrease in the stored sucrose concentration, resulting in up to doubled total sugar concentrations in harvested juice (Wu and Birch 2007). The transformed sugarcane lines with enhanced sugar accumulation also showed increased photosynthesis, sucrose transport and sink strength, indicating a possible feedback signal for the production, translocation, and storage of sucrose. This same group transformed sugarcane with a vacuole-targeted trehalulose synthase gene modified from the gene in *Pseusomonas mesoacididophila* MX-45 to obtain plants with mature stem juice concentrations of trehalulose reaching about 600 mM (Hamerli and Birch 2011). These plants retained vigor and trehalulose production over multiple vegetative generations under glasshouse and field conditions.

Using sugarcane as a biofactory is an exciting area of research that could have a tremendous impact on the sugarcane industries around the globe. However, many hurdles must be overcome before this can become a commercial reality. To be competitive with commercial protein production systems that use corn, alfalfa, rice, or tobacco, workers will need to achieve much higher levels of protein expression in the transgenic sugarcane stalk. This will require the continued identification of new promoters, development of viral vectors, and overcoming both transcriptional and posttranscriptional gene silencing. Furthermore, the details of how best to extract and purify proteins at large scales from plants are not trivial. Experience in this area, even with large companies, is scarce, particularly for sugarcane.

Biotechnology will not replace conventional sugarcane breeding, but it is another tool that can be used in conjunction with marker-assisted selection and traditional breeding for the continued improvement of sugarcane.

References

Aitken KS, Li J, Wang L, Qing C, Fan YH, Jackson P (2006) Characterization of intergeneric hybrids of *Erianthus rockii* and *Saccharum* using molecular markers. Genet Resour Crop Evol 54:1395–1405

Alexander AG (1985) The energy cane alternative. Elsevier, Amsterdam

Alix K, Baurens FC, Paulet JC, Glaszmann JC, D'Hont A (1998) Isolation and characterization of a satellite DNA family in the Saccharum complex. Genome 41:854–864

Alix K, Paulet JC, Glaszmann JC, D'Hont A (1999) Inter-Alu like species-specific sequences in the Saccharum complex. Theor Appl Genet 6:962–968

Amalraj VA, Balakrishnan R, Jebadhas AW, Balasundaram N (2006) Constituting a core collection of *Saccharum spontaneum* L. and comparison of three stratified random sampling procedures. Genet Resour Crop Evolut 53:1563–1572

Anonymous (1945) A newly released cane: some notes on NCo310. South Afr Sugar J 30:91

Arumuganathan K, Earle E (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9(3):208–218
Artschwager E, Brandes EW (1958) Sugarcane (Saccharum officinarum L.). US, Department of Agriculture
Balakrishnan R, Nair NV, Sreenivasan TV (2000) A method for establishing a core collection of *Saccharum officinarum* L. germplasm based on quantitative-morphological data. Genet Resour Crop Evolut 47:1–9
Barber CA (1920) The origin of the sugar cane International Sugar Journal 22:249–251
Bennett MD Leitch IJ (2003) Angiosperm DNA C-values Database (release 4.0, Jan 2003) http://wwwrbgkeworguk/cval/homepage.html
Berding N, Roach BT (1987) Germplasm collection, maintenance, and use. In: Heinz DJ (ed) Sugarcane improvement through breeding. Elsevier, Amsterdam, pp 143–210
Blume H (1985) Geography of sugar cane: environmental, structural and economical aspects of cane sugar production. In: Blume H (ed) Geography of sugar cane. Verlag Dr. Albert Bartens, Berlin, pp 21–36
Brandes EW (1956) Origin, dispersal and use in breeding of the Melanesian garden sugarcanes and their derivatives, *Saccharum officinarum* L. Proc Int Soc Sugar Cane Technol 9:709–750
Brandes EW (1958) Origin, classification and characteristics. In: Artschwager E, Brandes EW (eds) Sugarcane (*Saccharum officinarum* L.). U.S. Department of Agriculture Handbook 122, Washington, DC, pp 1–35
Bremer G (1923) A cytological investigation of some species and species-hybrids of the genus *Saccharum*. Genetica 5:273–326
Bremer G (1961) Problems in breeding and cytology of sugar cane. 4. Origin of increase of chromosome number in species hybrids of *Saccharum*. Euphytica 10(59–78):325–342
Brown JS, Schnell RJ, Power EJ, Douglas SL, Kuhn DN (2007) Analysis of clonal germplasm from five Saccharum species: *S. barberi*, *S. robustum*, *S. officinarum*, *S. sinense* and *S. spontaneum*. A study of inter- and intra species relationships using microsatellite markers. Genet Resour Crop Evolut 54:627–648
Brown JS, Schnell RJ, Tai PYP, Miller JD (2002) Phenotypic evaluation of *Saccharum barberi*, *S. robustum*, and *S. sinense* Germplasm from the Miami, Fl, USA world collection. Sugar Cane Int Sept–Oct:3–16
Bullard MJ, Heath MC, Nixon PM (1995) Shoot growth, radiation interception and dry matter production and partitioning during the establishment pase of *Miscnthus sinensis* "Giganteus" grown at two densities in the UK. Ann Appl Biol 126:94–102
Brumbley SM, Petrasovits L, Purnell M, O'Shea MG, Geijskes J, Lakshmanan P, Smith GR, Nielsen LK (2002) Application of biotechnology for future sugar industry diversification. Proc Aust Soc Sugar Cane Technol 24:40–46
Burnquist WL, Sorrells ME, Tanksley S (1992) Characterization of genetic variability in *Saccharum* germplasm by means of restriction fragment length polymorphism (RFLP) analysis. Proc Int Soc Sugar Cane Technol 21:355–365
Clayton WD (1972a) The awned genera of Andropogoneae. Studies in the Gramneae: 31. Kew Bull 27(3):457–454
Clayton WD (1972b) The awnless genera of Andropogoneae. Studies in the Gramneae: 33. Kew Bull 28(1):49–58
Clayton WD, Renvoize SA (1986) Genera *Graminum*—Grasses of the world. Kew Bull Additional Series 13:1–389
Clifton-Brown JC, Lewandowski I (2000) Over-wintering problems of new established Miscanthus plantations can be overcome by identifying genotypes with improved rhizome cold tolerance. New Phytol 148:287–294
Cuadrado A, Acevedo R, Dias M, de la Espina S, Jouve N, de la Torre C (2004) Genome remodelling in three modern *S. officinarum* × *S. spontaneum* sugarcane cultivars. J Exp Bot 55:847–854
D'Hont A (2005) Unravelling the genome structure of polyploids using FISH and GISH; examples of sugarcane and banana. Cytogenet Genome Res 109(1–3):27–33

D'Hont A, Grivet L, Feldmann P, Rao PS, Berding N, Glaszmann JC (1996) Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharun* spp.) by molecular cytogenetics. Mol Gen Genet 250:405–413

D'Hont A, Lu YH, Feldmann P, Glaszmann JC (1993) Cytoplasmic diversity in sugarcane revealed by heterologous probes. Sugar Cane 1:12–15

D'Hont A, Paulet F, Glaszmann JC (2002) Oligoclonal interspecific origin of 'North Indian' and 'Chinese' sugarcanes. Chromosome Res 10:253–262

D'Hont A, Rao PS, Feldmann P, Grivet L, Islam-Faridi N, Taylor P, Glaszmann JC (1995) Identification and characterization of sugarcane intergeneric hybrids, *Saccharum*-officinarum x *Erianthus-arundinaceus* with molecular markers and DNA in-situ hybridization. Theor Appl Genet 91:320–326

D'Hont A, Souza GM, Menossi M, Vincentz M, Van Sluys MA, Glaszmann JC, Ulian EC (2008) Sugarcane: a major source of sweetness, alcohol, and bio-energy. In: Moore PH, Ming R (eds) Genomics of tropical crop plants. Springer, New York, pp 483–513

Daniels J, Daniels C (1975) Geographical, historical and cultural aspect of the origin of the Indian and Chinese sugarcanes *S. barberi* and *S. sinense*. Int Soc Sugar Cane Technol Sugarcane Breed Newsl 36:4–23

Daniels J, Roach BT (1987) Taxonomy and evolution in sugarcane. In: Heinz DJ (ed) Sugarcane improvement through breeding. Elsevier, Amsterdam, pp 7–84

Daniels J, Roach BT, Daniels C, Paton N (1991) The taxonomic status of *Saccharum barberi* Jesweit and *S. sinense* Roxb. Sugar Cane 3:11–16

Daniels J, Smith P, Paton N, Williams C (1975) The origin of the genus *Saccharum*. Int Soc Sugar Cane Technol Sugarcane Breed Newsl 36:24–39

Deerr N (1921) Cane sugar, 2nd edn. Norma Roger, London

Deerr N (1949) The history of sugar, vol I. Chapman and Hall, London

Dohleman FG, Heaton EA, Leakey ADB, Long SP (2009) Does greater leaf-level photosynthesis explain the larger solar energy conversion efficiency of *Miscanthus* relative to switchgrass? Plant. Cell Environ 32(11):1525–1537

Edgerton CW (1958) Sugarcane and its diseases. Louisiana State Univ. Press, Baton Rouge, pp 43–61

Ethirajan AS (1987) Sugarcane hybridization techniques. In: Anonymous (eds) Copersucar International Sugarcane Breeding Workshop. Copersucar, Piracicaba, Brazil, pp 129–148

Gasson MJ, Kitamura Y, McLauchlan WR, Narbad A, Parr AJ, Parsons ELH, Payne J, Rhodes MJC, Walton NJ (1998) Metabolism of ferulic acid to vanillin. A bacterial gene of the enoyl-SCoA hydratase/isomerase superfamily encodes an enzyme for the hydration and cleavage of a hydroxycinnamic acid SCoA thioester. J Biol Chem 273:4163–4170

Glaszmann JC, Lu YH, Lanaud C (1990) Variation of nuclear ribosomal DNA in sugarcane. J Genet Breed 44:191–198

Gnanasambandam A, Birch RG (2004) Efficient developmental mis-targeting by the sporamin NTPP vacuolar signal to plastids in young leaves of sugarcane and Arabidopsis. Plant Cell Rep 24:435–447

Goldemberg J (2008) The Brazilian biofuels industry. Biotechnol Biofuels 1:6

Grivet L, Glaszmann JC, D'Hont A (2006) Molecular evidences for sugarcane evolution and domestication. In: Motley T, Zerega N, Cross H (eds) Darwin's harvest. New approaches to the origins, evolution, and conservation of crops. Columbia University Press, New York, pp 49–66

Grivet L, D'Hont A, Roques PD, Feldmann P, Lanaud C, Glaszmann JC (1996) RFLP mapping in cultivated sugarcane (*Saccharum* spp.): genome organization in a highly polyploid and aneuploid interspecific hybrid. Genetics 142:987–1000

Hamerli D, Birch R (2011) Transgenic expression of trehalulose synthase results in high concentrations of the sucrose isomer trehalulose in mature stems of field-grown sugarcane. Plant Biotechnol J 9(1):32–37

Heinz DJ, Osgood RV, Moore P (1994) Sugarcane. Encyclopedia of agricultural science, vol 4. Academic, San Diego, pp 225–238

Hoarau JY, Offmann B, D'Hont A, Risterucci AM, Roques D, Glaszmann JC, Grivet L (2001) Genetic dissection of a modern cultivar (*Saccharum* spp.). I. Genome mapping with AFLP. Theor Appl Genet 103:84–97

Hodkinson TR, Chase MC, Lledó M, Salamin N, Renvoize SA (2002) Phylogenetics of *Miscanthus*, *Saccharum* and related genera (Saccharinae, Andropogoneae, Poaceae) based on DNA sequences from ITS nuclear ribosomal DNA and plastid *trnL* intron and *trnL-F* intergenic spacers. J Plant Res 115:381–392

Holland-Moritz P (2003) Sugar cane-derived therapeutic proteins avoid contamination issues. Drug Discov Dev 6:27

Irvine JE (1999) Saccharum species as horticultural classes. Theor Appl Genet 98:186–194

Jannoo N, Grivet L, Dookun A, D'Hont A, Glaszmann JC (1999a) Linkage disequilibrium among modern sugarcane cultivars. Theor Appl Genet 99:1053–1060

Jannoo N, Grivet L, Seguin F, Paulet R, Domaingue PS, Rao A, Dookun A, D'Hont A, Glaszmann JC (1999b) Molecular investigation of the genetic base of sugarcane cultivars. Theor Appl Genet 99:171–184

Jeswiet J (1927) World collection of Saccharum. Proc Int Soc Sugar Cane Technol 2:137–139

Jeswiet J (1930) Proceedings of the International Society of Sugar cane Technologists Soerabaia

Kennedy AJ, Rao PS (2000) Handbook 2000. West Indies Central Sugar Cane Breeding Station, St. George, Barbados, pp 1–10

Lima MLA, Garcia AAF, Oliveira KM, Matsuoka S, Arizono H, de Souza CL Jr, de Souza AP (2002) Analysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugar cane (*Saccharum* spp.). Theor Appl Genet 104:30–38

Lu YH, Dhont A, Paulet F, Grivet L, Arnaud M, Glaszmann JC (1994a) Relationships among ancestral species of sugarcane revealed with RFLP using single copy maize nuclear probes. Euphytica 78:7–8

Lu YH, Dhont A, Paulet F, Grivet L, Arnaud M, Glaszmann JC (1994b) Molecular diversity and genome structure in modern sugarcane cultivars. Euphytica 78:217–226

Machado GR, Da Silva WM, Irvine J (1987) Sugarcane breeding in Brazil: the Copersucar program. In: Anonomous (eds) Copersucar International Sugarcane Breeding Workshop. Copersucar, Brazil, pp 216–247

McQualter RB, Fong Chong B, O'Shea MG, Meyer K, Van Dyk DE, Viitanen PV, Brumbley SM (2005) Initial evaluation of sugarcane as a production platform for p-hydroxybenzoic acid. Plant Biotechnol J 3(1):29–41

Mukherjee SK (1954) Revision of the genus *Saccharum* Linn. Bull Bot Soc Bengal 89:143–148

Mukherjee SK (1957) Origin and distribution of *Saccharum*. Bot Gaz 119:55–56

Nair NV, Nair S, Sreenivasan TV, Mohan M (1999) Analysis of genetic diversity and phylogeny in Saccharum and related genera using RAPD markers. Genet Res Crop Evolut 46:73–79

Nuss KJ, Brett PGC (1995) The release of cultivar NCo310 in 1945 and its impact on the sugar industry. Proc South Afr Sugar Technol Assoc 69:3–8

Panje RR, Babu CN (1960) Studies in *Saccharum spontaneum*. Distribution and geographical association of chromosome numbers. Cytologia 25:152–172

Parthasarathy N (1948) Origin of Noble Sugar-Canes (*Saccharum officinarum*). Nature 161:606–608

Paton N, Daniels J, Smith P (1978) A study of *S. sinense* Roxb. and *S. barberi* Jesw. Int Soc Sugarcane Technol Sugarcane Breed Newsl 41:33–50

Peoples OP, Sinskey AJ (1989) Poly-beta-hydroxybutyrate (PHB) biosynthesis in Alcaligenes eutrophyus H16. Characterization of the genes encoding beta-ketothiolase and acetoacetyl-CoA reductase. J Biol Chem 264:15293–15297

Peoples OP, Sinskey AJ (1989) Poly-beta-hydroxybutyrate (PHB) biosynthesis in Alcaligenes eutrophyus H16. Identification and characterization of the PHB polymerase gene (phbC). J Biol Chem 264:15298–15303

Piperidis N, Chen J-W, Deng H-H, Wang LP, Jackson P, Piperidis G (2010a) GISH characterization of *Erianthus arundinaceus* chromosomes in three generations of sugarcane intergeneric hybrids. Genome 53:331–336

Piperidis G, Christopher MJ, Carroll BJ, Berding N, D'Hont A (2000) Molecular contribution to selection of intergeneric hybrids between sugarcane and the wild species *Erianthus arundinaceus*. Genome 43:1033–1037

Piperidis G, Piperidis N, D'Hont A (2010b) Molecular cytogenetic investigation of chromosome composition and transmission in sugarcane. Mol Genet Genomics 284:65–73

Price HJ, Dillon SL, Hodnett G, Rooney WL, Ross L, Johnston JS (2005) Genome evolution in the genus Sorghum (Poaceae). Ann Bot 95:219–227

Price S (1957) Cytological studies in Saccharum and allied genera II. Chromosome numbers in interspecific hybrids. Bot Gaz 118:146–159

Price S (1963) Cytogenetics of modern sugar canes. Econ Bot 17:97–105

Price S (1965) Interspecific hybridization in sugarcane breeding. Proc Int Soc Sugar Cane Technol 12:1021–1026

Roach BT (1989) Origin and improvement of the genetic base of sugarcane. Proc Aust Soc Sugar Cane Technol 11:34–47

Roach BT, Daniels J (1987) The *Saccharum* complex and the genus *Saccharum*. In: Anomous (eds) Copersucar Int. Sugarcane Breeding Workshop. Copersucar, Brazil, pp 1–33

Rossi G, da Silva W, Irvine J (1987) Sugarcane breeding in Brazil: the Copersucar program. In: Anomous (eds.) Copersucar International Sugarcane Breeding Workshop. Copersucar, Brazil, pp 217–232

Schubert P, Steinbuchel A, Schlegel HG (1988) Cloning of the Alcaligenes eutrophus genes for synthesis of poly-beta-hydroxybutyric acid (PHB) and synthesis of PHB in Escherichia coli. J Bacteriol 170:5837–5847

Siebert M, Sommer S, Li S, Wang Z, Severin K, Heide L (1996) Genetic engineering of plant secondary metabolism. Accumulation of 4-hydroxybenzoate glucosides as a result of the expression of the bacterial ubiC gene in tobacco. Plant Physiol 112:811–819

Simmonds NW (1976) Sugarcanes. In: Simmonds NW (ed) Evolution of crop plants. Longman Group Limited, London, pp 104–108

Sobral BWS, Braga DPV, Lahood ES, Keim P (1994) Phylogenetic analysis of chloroplast restriction enzyme site mutations in the Saccharinae Griseb. Subtribe of the Andropogoneae Dumort. Tribe. Theor Appl Genet 87:843–853

Spangler R, Zaitchik B, Russo E, Kellogg E (1999) Andropogoneae evolution and generic limits in Sorghum (Poaceae) using ndhF sequences. Syst Bot 24:267–281

Sreenivasan TV, Ahloowalia BS, Heinz DJ (1987) Cytogenetics. In: Heinz DJ (ed) Sugarcane improvement through breeding. Elsevier, Amsterdam, pp 211–253

Stevenson GC (1965) Genetics and breeding of sugar cane. Longman, London

Tai PYP, Miller JD (2001) A core collection for *Saccharum spontaneum* L. from the World Collection of sugarcane. Crop Sci 41:879–885

Tai PYP, Miller JD (2002) Germplasm diversity among four sugarcane species for sugar composition. Crop Sci 42:958–964

Tew TL (1987) New varieties pp. In: Heinz DJ (ed) Sugarcane improvement through breeding. Elsevier, Amsterdam, pp 559–594

Tew TL (2003) World sugarcane variety census – Year 2000. Sugar Cane International March/April 2003:12–18

Tew TL, Cobill RM (2008) Genetic improvement of sugarcane (*Saccharum* spp.) as an energy crop. In: Vermerris W (ed) Genetic improvement of bioenergy crops. Springer Science LLC, New York

Waclawovsky AJ, Sato PM, Lembke CG, Moore PH, Souza GM (2010) Sugarcane for bioenergy production: an assessment of yield and regulation of sucrose content. Plant Biotechnol J 8:1–14

Walker DIT (1987) Manipulating the genetic base of sugarcane. In: Anonymous (eds.) Copersucar International Sugarcane Breeding Workshop. Copersucar, Piracicaba, Brazil, pp 321–334

Wang ML, Goldstein C, Su W, Moore PH, Albert H (2005) Production of biologically active GM-CSF in sugarcane: a secure biofactory. Transgenic Res 114:167–178

Whalen MD (1991) Taxanomy of Saccharum (Poaceae). Baileya 23:109–125
Wu L, Birch RG (2007) Doubled sugar content in sugarcane plants modified to produce a sucrose isomer. Appl Environ Microbiol 71:1581–1590
Xavier RM (2007) The Brazilian ethanol experience. Competetive Enterprise Institute, Washington, DC http://www.cei.org/pdf/5774.pdf
Yuan JS, Tiller KH, Al-Ahmad H, Stewart NR, Stewart CN Jr (2008) Plants to power: bioenergy to fuel the future. Trends Plant Sci 13:421–429

Chapter 4
The Gene Pool of *Miscanthus* Species and Its Improvement

Erik J. Sacks, John A. Juvik, Qi Lin, J. Ryan Stewart, and Toshihiko Yamada

Abstract For more than a thousand years, people have used *Miscanthus* from wild stands or managed landscapes, to feed their livestock, roof their homes, make paper, dye possessions, and beautify their gardens. In recent decades there has been a call to develop *Miscanthus* into a fully domesticated biomass crop for sustainable renewable energy needs. *Miscanthus* is broadly distributed throughout eastern Asia and the Pacific islands, ranging from southern Siberia to tropical Polynesia, with a current center of diversity in temperate northern latitudes. Adaptation to cold and temperate environments is a distinctive feature of *Miscanthus* relative to other Saccharinae, facilitating its potential to become an important biomass crop in Europe and the USA. Auto- and allopolyploidy have played a role in the evolution of *Miscanthus* and polyploidy will likely be of central importance for the development and improvement of this crop. Variation for flowering time, including short-day flower induction, will permit plant breeders to optimize local adaptation and biomass-yield of *Miscanthus*, just as they have done for maize, sorghum and sugarcane. Germplasm collections that are representative of the genus and publicly available need to be established and characterized. Questions of taxonomy, origins, and evolution need attention from the research community. A multidisciplinary approach that includes population genetics, cytogenetics, molecular genetics, and genomics

E.J. Sacks (✉) • J.A. Juvik • J.R. Stewart
Department of Crop Sciences, University of Illinois, 1101 Institute for Genomic Biology, 1206W. Gregory Avenue, Urbana, IL 61801, USA
e-mail: esacks@illinois.edu; juvik@illinois.edu; rstewart@illinois.edu

Q. Lin
Institute of Botany, Chinese Academy of Sciences, Beijing, China
e-mail: linqi@ibcas.ac.cn

T. Yamada
Field Science Center for Northern Biosphere, Hokkaido University, Sapporo, Hokkaido, Japan
e-mail: yamada@fsc.hokudai.ac.jp

A.H. Paterson (ed.), *Genomics of the Saccharinae*, Plant Genetics and Genomics: Crops and Models 11, DOI 10.1007/978-1-4419-5947-8_4,

will be needed to rapidly increase our knowledge of the *Miscanthus* gene pool, which will facilitate the development of improved cultivars.

Keywords *Miscanthus* • *Saccharum* • Miscanes • Taxonomy • Centers of origin • Ploidy • Interspecific hybridization • Traditional uses • Breeding

1 Singular Properties of the Genus and Its Members

Miscanthus is a perennial, warm-season C_4 grass. Some genotypes are highly productive in temperate environments. Depending on environment and genetics, high yields (10–41 dry t ha^{-1}) of cultivated *Miscanthus* can be achieved with low inputs (Clifton-Brown et al. 2004; Heaton et al. 2004, 2008; Xi 2000). Stands of cultivated *Miscanthus* can remain productive for more than 10 years (Clifton-Brown and Jones 2001). Thus, over the last 20 years in Europe and more recently in the USA, *Miscanthus* has been the focus of research to determine its potential as a sustainable biomass crop for meeting renewable-energy needs (Clifton-Brown et al. 2004; Scurlock 1998). Self-incompatibility (Hirayoshi et al. 1955) and adaptation to diverse environments has resulted in great genetic diversity among and within natural *Miscanthus* populations but little of this has yet been used for crop development or improvement.

Miscanthus has a broad natural geographic distribution in eastern Asia and throughout the Pacific islands, from about 50°N in southern Siberia to 22°S (Hodkinson et al. 1997, 2002a). A key distinctive feature of *Miscanthus* relative to other members of the *Saccharinae* is that centers of *Miscanthus* genetic diversity are in northern temperate latitudes. Thus, adaptation to temperate climates is a feature of many *Miscanthus* populations, making the genus especially attractive for development of a perennial biomass crop adapted to North America and Europe.

Adaptation of some *Miscanthus* populations to temperate environments can manifest as the ability to over-winter though USDA hardiness zone 3, and/or as spring regrowth preceeding other warm-season C_4 grasses by 2–4 weeks. Perhaps the most important and fundamental adaptation of *Miscanthus* to temperate environments is the ability of some genotypes to maintain high photosynthetic rates at temperatures below 12 °C (Beale et al. 1996; Naidu et al. 2003; Wang et al. 2008). In contrast to *Miscanthus*, CO_2 assimilation in its *Saccharinae* relatives, sugarcane, maize, and sorghum, is greatly reduced at temperatures below 14 °C. Thus, *Miscanthus* can produce more biomass per season than maize in the Midwest US (Dohleman and Long 2009) and in southern England (Wang et al. 2008).

Miscanthus is closely related to *Saccharum officinarum* and its primary wild progenitor, *S. robustum* (Amalraj and Balasundaram 2006; Hodkinson et al. 2002a). Sugarcane improvement programs have produced fertile intergeneric progeny of *Miscanthus* and *Saccharum* (Chen 1993; Chen and Lo 1989; Li et al. 1951). Morphological distinctions between the genera are primarily the presence in *Miscanthus* of pedicels on both of the paired florets, with one pedicel longer than the other, whereas in *Saccharum* one floret of each pair is pedicellate and one is sessile.

The name *Miscanthus*, derived from the Greek words *mischos* for stalk and *anthos* for flower, refers to this distinctive anatomy. In addition, the tough rachis of *Miscanthus* contrasts with the fragile rachis of *Saccharum*. Cytogenetically, *Miscanthus* differs from *Saccharum* by having a basic chromosome number of $x=19$ in contrast to $x=10$, which predominates in the *Saccharinae*. How $x=19$ evolved has been a subject of considerable interest. Hirayoshi et al. (1959); Adati and Shiotani (1962); Grassl (1974) hypothesized that *Miscanthus* is an amphiploid derived from a cross between an n=10 and n=9 species. Recently, three research groups that developed genetic maps of *Miscanthus* came to the same conclusion that $x=19$ resulted from a whole genome duplication and the fusion of two chromosomes (Kim et al. 2012; Swaminathan et al. 2012; Xue-Feng et al. 2012). The chromosome fusion event appears to have occurred after the genome duplication (Xue-Feng et al. 2012), and thus $x=19$ was likely not the result of interspecific hybridization.

A group of African *Miscanthidium* species have been placed in *Miscanthus* by some workers but these differ from the eastern Asian species by having a basic chromosome number of $x=15$ and molecular marker profiles indicating that they are more closely related to *Saccharum narenga* and *Saccharum* (*Erianthus*) *contortum* (Amalraj and Balasundaram 2006; Hodkinson et al. 1997, 2002a). Similarly, the Himalayan-centered *Diandranthus* species have often been placed in *Miscanthus* but they have a chromosome number ($2n=40$), stamen number (2 in contrast to 3), and a molecular marker profile that indicates that they are more closely related to *Sorghum* than to the eastern Asian *Miscanthus* (Amalraj and Balasundaram 2006; Hodkinson et al. 1997, 2002; Shouliang and Renvoize 2006).

The genus *Miscanthus* can be divided into two sections (Lee 1993): sect. *miscanthus*, which includes *M. sinensis*, *M. floridulus*, *M. tinctorius*, *M. oligostachyus*, and *M. intermedius*, and sect. *triarrhena*, which includes *M. sacchariflorus* and the economically important *M. sacchariflorus var. lutarioriparius* (syn. *M. lutarioriparius* or *Triarrhena lutariariparia*). Distinguishing phenotypic characteristics of Sect. *miscanthus* are tufted habit, typically with short rhizomes, nodes on culms typically lacking buds except at the base of the culm, no adventitious roots, spikelets typically with awns though sometimes awnless, and leaves that are usually pilose on the abaxial surface but sometimes glabrous. In contrast, sect *triarrhena* is distinguished by a strongly rhizomatous and spreading habit, culms typically having buds at the nodes, adventitious roots, awnless spikelets, and leaves that are glabrous on the abaxial surface.

2 Centers of Diversity, Secondary and Tertiary Gene Pools, Germplasm Resources

2.1 Centers of Diversity

The primary center of *Miscanthus* diversity is in China, especially between ~25–35°N and ~110–120°E (Sun et al. 2010). In this region *M. sinensis*, *M. floridulus*, *M. sacchariflorus*, and *M. sacchariflorus* var. *lutarioriparius* are found. Individuals

in the Chinese populations are nearly always diploid though rare triploids or tetraploids are sometimes observed. Many local varieties are found in China, reflecting the wide distribution of the species and the broad range of climates, latitudes, altitudes and soil types to which *Miscanthus* has adapted.

Japan is an important secondary center of diversity. In contrast to China, where polyploid individuals are rarely observed, polyploidy has been important in the evolution of *Miscanthus* in Japan. In Japan there are three endemic species of subsect. *kariyasua* (Ibaragi and Ohashi 2004; Sun 2009): *M. tinctorius*, *M. oligostachyus*, and *M. intermedius*. The former two species are diploid but *M. intermedius* is a hexaploid with apparently multiple homologous copies of a single genome (Adati and Shiotani 1962). *M. sinensis* and *M. floridulus* are also found in Japan. Notably absent from Japan are reports of diploid *M. sacchariflorus* (Adati and Shiotani 1962; Hirayoshi et al. 1957).

However, an allopolyploid species derived from *M. sinensis* and *M. sacchariflorus* is found throughout much of Japan (Adati 1958a; Adati and Shiotani 1962; Hirayoshi et al. 1957). These allopolyploids are predominantly tetraploid, but less frequently triploid or pentaploid. They are highly rhizomatous and awnless, and have frequently been described in the literature as *M. sacchariflorus* (Ogi in Japanese) because of the morphological similarity to that species (Adati and Shiotani 1962; Hirayoshi et al. 1957; Hodkinson et al. 2002b). Similar polyploids have not been found in China. There is currently insufficient data to exclude the possibility that some of the polyploid Japanese Ogi are autotetraploid forms of *M. sacchariflorus* but the available data does not suggest that this is likely. Evidence for allopolyploidy of Japanese Ogi comes from studies of chromosome pairing, karyotype studies, estimates of genome size, and DNA sequence data. Adati and Shiotani (1962) observed that natural triploid Japanese Ogi produce an approximately complete set of bivalents with occasional trivalents during meiosis, and that an artificial triploid progeny, derived from diploid *M. sinensis*/tetraploid-Ogi, behaved similarly, indicating the presence of two different types of genomes, one of which was fully homologous to *M. sinensis* and the other of which was only partially homologous to *M. sinensis*. Adati and Shiotani (1962) discussed that the natural triploid plant was derived from a cross of *M. sacchariflorous* and *M. sinensis*. Further support for the presence of two types of genomes in Japanese Ogi comes from observations that tetraploids had only one pair of intercalary trabant chromosomes (SAT-chromosomes) and triploids had one or two SAT-chromosomes (Adati 1958a; Adati and Shiotani 1962; Lafferty and Lelley 1994; Linde-Laursen 1993). In contrast, autopolyploids within sect. *miscanthus* were observed to have a number of SAT-chromosomes equivalent to the number of genomes; diploids had two, triploids had three and hexaploids had six (Adati and Shiotani 1962). Rayburn et al. (2009) found that the genome size of diploid *M. sacchariflorus* was 4.5 pg, *M. sinensis* was 5.5 pg and an allotriploid was 7.0 pg, which is consistent with expectations. Lastly, by comparing nuclear (*nrITS*) and chloroplast (*trnL-F*) DNA sequences of a triploid Japanese Ogi with diploid *M. sinensis* and diploid *M. sacchariflorous*, Hodkinson et al. (2002c) found that the sequence variation was consistent with the conclusion that the triploid was an allopolyploid derived from *M. sacchariflorus* and *M. sinensis* and

that the female ancestor was *M. sacchariflorus*. Sequence data from Swaminathan et al. (2010) generally agreed with the data from Hodkinson et al. (2002c). Hodkinson and Renvoize (2001) formally defined the nothospecies, *M.* ×*giganteus*, as a hybrid of *M. sinensis* and *M. sacchariflorus*. Matumura (1998a) speculated *M.* ×*giganteus* to be the same taxon as *Miscanthus ogiformis* (Honda 1939).

The Korean peninsula, with its long north–south axis and associated range of environments, is another secondary center of *Miscanthus* diversity. Lee (1993) indicated that *M. sinensis* and *M. sacchariflorus* are present in Korea. Interestingly, Lee's (1993) descriptions suggest that both diploid *M. sacchariflorus* and *M. sinensis*/*M. sacchariflorus* allopolyploids may be present in Korea. An especially intriguing finding from an AFLP diversity study of Korean and Japanese *Miscanthus* was that one Korean plant identified as *M. sacchariflorus* was in an isolated phylogenetic position between the other Korean *M. sinensis* and *M. sacchariflorus* genotypes, which is what one would expect to see for an allopolypoid (Lledó et al. 2001). Unfortunately Lledó et al. (2001) did not report on the cytogenetic constitution of the materials in their molecular marker study. Further work is needed to clarify the range of the *M. sinensis*/*M. sacchariflorus* allopolyploids. Lee (1993) also reported the presence of two subsect. *kariyasua* species in Korea: *M. oligostachyus* var. *intermedius*, and an awned, tufted endemic species, *M. changii*. Ibaragi and Ohashi (2004) revised subsect. *kariyasua* and reclassified *M. oligostachyus* var. *intermedius* as *M. longiberbis* var. *longiberbis* and *M. changii* as *M. longiberbis* var. *changii*. However, neither of the Korean subsect. *kariyasua* classifications were supported by cytogenetic or molecular marker data. Nevertheless, the published reports taken together suggest that Korea may represent an important evolutionary link between *Miscanthus* populations on the Asian mainland and those on the nearby islands of Japan.

2.2 *M. sinensis*

In Asia, *M. sinensis* is the most broadly distributed *Miscanthus* species. It is found primarily in China, Korea, and Japan (Lee 1993; Lee 1964a, b; Shouliang and Renvoize 2006; Sun 2009; Sun et al. 2010). *M. sinensis* is also reported to be found through mainland Southeast Asia to Borneo and the Philippines (Inthakoun and Delang 2008; Lee 1964b; Newman et al. 2007). Thus, its northern range extends to USDA hardiness zone 4 in Liaoning and Jilin provinces in China, and Primorsky Krai, the Kuril Islands, and Sakhalin in Russia, while in the southernmost part of its range it is adapted to lowland and upland tropics. *M. sinensis* is found from sea level to 2,500 m, though most plants at high elevations are in the central and southern parts of the species' range. It is commonly found on mountain slopes, in open grasslands, on roadsides, and in open coastal areas (Lee 1964b; Shouliang and Renvoize 2006; Stewart et al. 2009; Sun et al. 2010). Some populations very near the ocean shore may be sources of tolerance to salt (Chiang et al. 2003a). Well-drained sites are typically preferred, though occasionally populations are found in wet soils near fresh water. Areas that are regularly but infrequently disturbed, such as by grazing,

mowing or burning, favor the development and maintenance of large populations by preventing trees and shrubs from outcompeting grassland plants (Iwanami 1969; Stewart et al. 2009). Molecular marker studies of *M. sinensis* have begun to characterize genetic variability within natural populations and groups of European ornamental cultivars, and though sampling so far has been limited, the results are consistent with the wide morphological diversity observed (Greef et al. 1997; Hodkinson et al. 2002b; Lledó et al. 2001).

A winter dormancy response to changing conditions in the autumn is typical of most *M. sinensis* populations from temperate environments. In contrast, populations from tropical and subtropical environments (e.g., Hainan, Guangxi, Guangdong, Yunnan, and Taiwan), as well as var. *condensatus* from coastal Japan, are typically evergreen. In fact, under favorable conditions, such as a warm greenhouse in IL, many northern-adapted *M. sinensis* genotypes will continue to grow actively. However, genotypes that are adapted to be evergreen in tropics or subtropics do not typically go dormant under field conditions in cold-temperate environments.

M. sinensis plants typically flower once per year, primarily towards the end of the growing season from August to November. Northern-adapted populations of *M. sinensis* are day-neutral (Clifton-Brown et al. 2008). Indeed, some *M. sinensis* genotypes can initiate flowers under 24 h of daylight. In contrast to the northern materials, short-day flower initiation is typical in *M. sinensis* populations from the southernmost Chinese provinces of Hainan, Guangxi and Yunnan, and presumably in the adjacent countries to the south. If grown in a common garden, northern genotypes typically flower earlier than southern genotypes, even if only day-neutral types are compared. Thus, flowering time may be regulated by response to growing degree days, photoperiod or both (Jensen 2009). Height typically ranges from 0.7 to 3 m (Lee 1993; Lee 1964a, b; Renvoize 2003; Shouliang and Renvoize 2006; Sun et al. 2010). Tall genotypes are typically more common in the south than in the north but this is related to flowering time, as each culm terminates in an inflorescence. Clifton-Brown et al. (2001b) also observed that *Miscanthus* height and yield are influenced by flowering time. Optimization of flowering time, especially by selecting for presence, absence or degree of response to day length has been a major theme of domestication in the *Saccharinae* and for crops in general. Thus, variation for photoperiodic control of flowering represents an important opportunity for *Miscanthus* improvement (Jensen 2009).

M. sinensis var. *condensatus* from coastal areas throughout Japan and from the island of Taiwan is potentially an important resource for breeding improved cultivars for biomass yield and for studies of *Miscanthus* genetics. Indeed, several cultivars of var. *condensatus*, 'Imazo', 'Yoreba', 'Hiroba', 'Boukou', and 'Nogura' were developed by farmers and used as forages on the Izu Islands of Japan (Matumura 1998b). Genotypes of var. *condensatus* are typically robust, tall (2–4 m), late flowering and have wide leaves. The ornamental cultivars 'Cabaret', 'Cosmopolitan', and 'Emerald Shadow' are examples of var. *condensatus* that are common in the USA. Though var. *condensatus* possesses traits of value for breeding high-biomass cultivars, their tendency to stay green past initial frosts makes them less hardy in

cold continental climates than varieties with a strong dormancy response. For example, 'Cabaret' plants grown in the field in central Illinois, USA (USDA hardiness zone 6) remain green and undamaged in the late autumn through −4 °C but the leaves freeze out when temperatures drop below −6 °C. In the southernmost part of its natural range, populations of var. *condensatus* have been reported to be self-compatible or apomictic (Chiang et al. 2003b; Chou et al. 2000). A self-compatible *Miscanthus* would be a useful tool for genetic studies, especially for identifying phenotypic effects of low-frequency recessive genes (e.g., nonshattering) which seldom achieve homozygosity in self-incompatible populations. In addition to being useful for studying naturally occurring recessive genes, self-compatibility would facilitate the use of mutagenesis strategies for understanding gene function. On the other hand, apomixis could enable plant breeders to produce uniform, heterozygous, clonal varieties from seed.

For over a century, *M. sinensis* cultivars have been grown as ornamentals in Europe and the USA (Darke 1994, 2007; Deuter and Abraham 1998). More than 85 cultivars are currently sold in the US nursery trade. Anecdotally, among these ornamental *M. sinensis* cultivars, height ranges from 0.3 to 3 m and the start of flowering ranges from June to November. Thus, the *M. sinensis* germplasm currently in Europe and the USA encompasses considerable genetic diversity that can be used to develop high-biomass *Miscanthus* cultivars and for studying the inheritance of key traits.

2.3 *M. floridulus*

M. floridulus is morphologically similar to *M. sinensis* but is primarily adapted to tropical and subtropical environments. *M. floridulus* is a robust, tall (1–5 m), evergreen species (Lee 1964a, Lee 1964c; Shouliang and Renvoize 2006; Sun et al. 2010). In addition to height, characteristics that typically distinguish *M. floridulus* from *M. sinensis* (Lee 1964a, b, c Shouliang and Renvoize 2006; Sun et al. 2010) include large panicles (axis 16–45 cm) with numerous racemes (20–100) and spikelets that are short (2.5–4 mm), in contrast to short panicles (axis 3–22 cm) with several racemes (4–40) and spikelets that are long (4–7 mm). The abaxial leaf surface of *M. floridulus* is typically glabrous but in *M. sinensis* it is usually pilose. Also, the lower racemes of *M. floridulus* are longer than the upper ones, often giving the inflorescence the overall shape of a chef's knife (broad at the base and narrow at the top), which is a distinct appearance from the more uniformly long racemes of *M. sinensis*. In China, *M. floridulus* has a unique bimodal distribution for flowering time, flowering at the beginning and again at the end of the summer rainy season. Though typical forms of *M. floridulus* and *M. sinensis* can be readily distinguished from each other, the distributions of phenotypic characteristics of both species overlap (Sun 2009; Sun et al. 2010). Moreover, interbreeding between the two species can occur where they occupy the same geographical area. Thus, reports of *M. sinensis* in tropical areas that are based solely on morphology should be interpreted

with caution. Molecular marker studies will be useful for clarifying the southernmost distribution of *M. sinensis* (Scally et al. 2001).

The geographic range of *M. floridulus* extends from ~35°N in southern coastal Japan and in Anhui and Hubei provinces in China (USDA hardiness zone 8), throughout tropical Southeast Asia to Micronesia, New Guinea and Polynesia (Christophersen 1935; Florence and Lorence 1997; Haberle 2007; Lee 1964a, c). In much of its range, *M. floridulus* is found primarily in coastal lowlands but in New Guinea it is found in from 1,000 to 2,800 m (Haberle 2007; Lee 1964a; Reeder 1948). The presence of *M. floridulus* on two large land masses that are at opposite ends of the species' north–south range leads to a fundamental question of its origin: Did *M. floridulus* originate in mainland Southeast Asia and spread southward or did it originate in New Guinea and spread northward? Lee (1964a) considered *M. floridulus* the most ancestral of the *Miscanthus* lineage and this is supported currently by limited molecular marker data (Chiang et al. 2003b) though in other studies such a relationship is not evident (Hodkinson et al. 2002a). In a common-garden experiment, Chou et al. (2001) found that Taiwanese *M. floridulus* that grew well at 1,000 m did not survive a colder environment at 2,600 m, whereas upland *M. sinensis* var. *transmorrisonensis* grew well at both test sites, indicating that adaptation to cooler environments is a derived trait and that adaptation to warm environments is the ancestral state for *Miscanthus*. A New Guinea origin of *Miscanthus* is an especially intriguing hypothesis because some of its closest relatives (e.g., *S. robustum* and *S. officinarum*) possibly also originated in New Guinea (Amalraj and Balasundaram 2006; Hodkinson et al. 2002a; Ramdoyal and Badaloo 2002). However, Grassl (1974) argued that *Miscanthus* and *Saccharum* originated in Asia. Molecular genetics studies should shed light on the origins of *Miscanthus* but to-date, sampling of germplasm, especially of *M. floridulus*, has been too limited to draw conclusions. Though the origin of *Miscanthus* is currently uncertain, an answer to the question would impact strategies for conserving germplasm and for using it to develop biomass cultivars.

2.4 *M. sacchariflorus*

M. sacchariflorus is a strongly rhizomatous, temperate-adapted species, with the most northerly range in the genus. It is found in southern-Siberia, China and Korea. With rare exception, it is nearly always diploid. The type region of *M. sacchariflorus* is the Amur River watershed, which is USDA hardiness zone 3 (Hictchcock 1971; Lee 1993; Maximowicz 1859). The southern range of *M. sacchariflorus* extends to the Yangzi River watershed at ~28°N in northern Jiangxi and Hunan provinces in China (USDA hardiness zone 9). In contrast to other *Miscanthus* species, it is commonly, though not exclusively, found in moist soils along the banks of rivers and lakes (Lee 1993; Shouliang and Renvoize 2006; Sun et al. 2010). In contrast to *M. sinensis* and *M. floridulus*, stems of *M. sacchariflorus* can be hollow, which may be an adaptation to flooded soils. The typical form of the species is from 0.7 to

2.5 m tall with culms 5–10 mm in diameter (Lee 1993; Shouliang and Renvoize 2006; Sun et al. 2010). Taller genotypes are more prevalent in the south than in the north. The typical form of the species flowers early in late summer or early autumn, apparently is day-neutral for flower initiation, and has a strong dormancy response in the early autumn. In fact, northern-adapted *M. sacchariflorus* can go dormant in the autumn even when grown in a warm greenhouse, and such dormant plants can regrow in the spring. *M. sacchariflorus* 'Robustus', which can reach 2.4 m and is very cold-hardy, has been included in the living collections of many botanical gardens in Europe and the USA, and has been sold in the nursery trade. Cultivation of *M. sacchariflorus* 'Robustus', in Europe may date back to the late nineteenth century from the collections by Maximowicz (M. Deuter, personal communication).

M. sacchariflorus var. *lutarioriparius* is a tall (3–7 m), thick-stemmed (8–25 mm), strongly rhizomatous, highly spreading, and vigorous plant that is found in seasonally flooded riparian sites in the Yangtze River watershed east of 110°E and from 28° to 35°N (Fig. 4.1) (Chen and Renvoize 2005; Liu and Yu 2004; Sun 2009; Sun et al. 2010; Xi 2000). Unlike the typical form of the species, *M. sacchariflorus* var. *lutarioriparius* flowers late, in October, and culms can have more than twice as many nodes, indicating that it is likely a short-day plant. It is cultivated on large areas, in near monocultures along the edges of lakes and rivers. Yearly harvests of dry stems are stored for commercial production of paper (Xi 2000). Xi (2000) reports yields in China of 22–41 dt ha^{-1}. In contrast to most other *Miscanthus*, *M. sacchariflorus* var. *lutarioriparius* sheds its older lower leaves at the nodes, leaving behind clean culms. Leaf loss in *M. sacchariflorus* var. *lutarioriparius* may have evolved as an adaptation to reduce drag during seasonal flooding. Leaf loss is potentially a quite useful trait for improving the quality of biomass cultivars because leaves contain a greater proportion of minerals relative to cellulose and lignin than culms.

2.5 *M. ×giganteus*

Hodkinson and Renvoize (2001) based their description of *M. ×giganteus* on a sterile triploid individual that was imported from Yokohama, Japan to Denmark in the 1930s by 'Aksel Olson' (Deuter and Abraham 1998; Greef and Deuter 1993; Greef et al. 1997; Linde-Laursen 1993; Nielsen 1990). This genotype is high-yielding (14–30 dt $ha^{-1)}$, tall (3–4+ m) and broadly adapted to many agricultural regions of Europe and the USA (Clifton-Brown et al. 2001b, 2004; Heaton et al. 2004, 2008; Pyter et al. 2007). A key part of its broad adaptation to temperate environments is its ability to maintain high rates of photosynthesis at low temperature (Beale et al. 1996; Dohleman and Long 2009; Naidu et al. 2003; Wang et al. 2008). High yield and height are at least partly achieved by late flowering, which might be attributed to a moderate, short-day flowering response. In Germany and the UK, it does not flower before the first autumn frost in most years, yet a dormancy response is typically observed (Clifton-Brown et al. 2001b; Pyter et al. 2007), demonstrating at least a partial decoupling of dormancy and flowering in this genotype. It is clonally

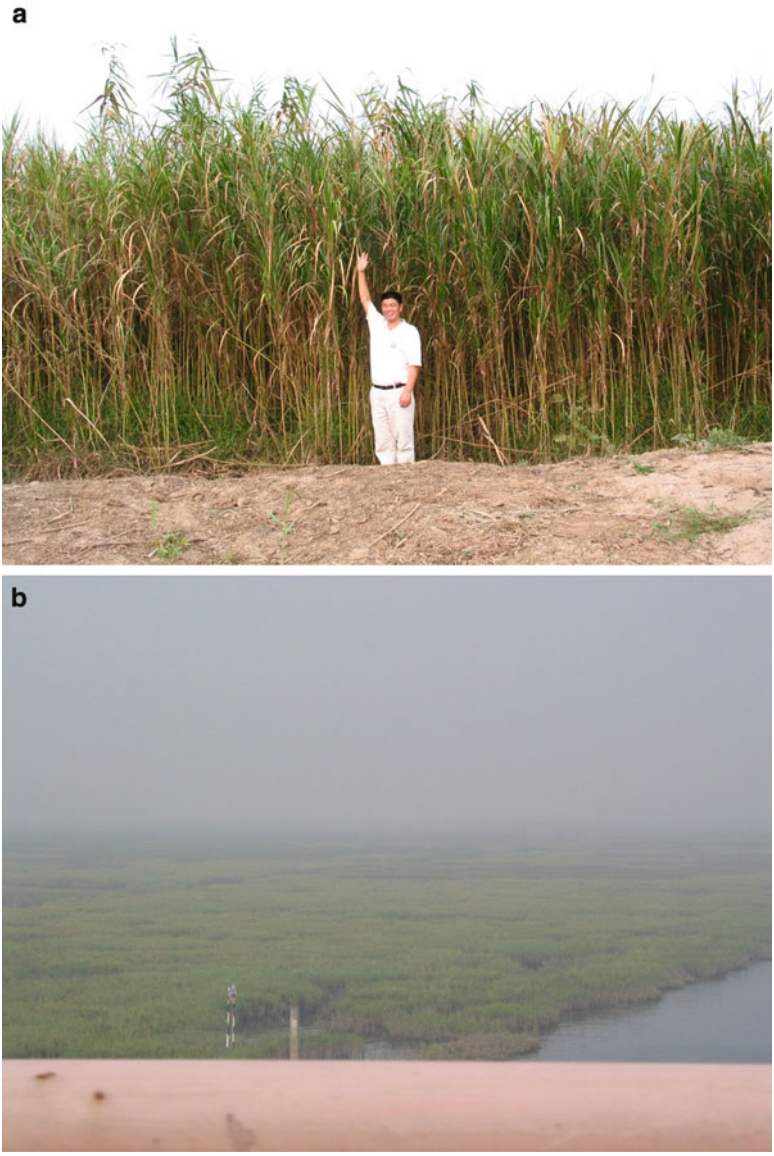

Fig. 4.1 *Miscanthus sacchariflorus* var. *lutarioriparius* grown in Hunan, China as a feedstock for paper production

propagated, typically from rhizome cuttings, generating mature stands by years 3–4 and providing yields which can be maintained for 20 or more years of production (Lewandowski et al. 2000). Once established, the stands are so dense that there is no need for weed control. It has greater water use efficiency than corn (Beale et al. 1999) and because of its capacity to reallocate minerals and other nutrients to the

underground rhizomes in the fall (Christian and Haase 2001), has extremely high nitrogen use efficiency (Beale and Long 1997). Molecular marker data indicate that there is currently one predominant *M.* ×*giganteus* genotype in Europe and the USA (Greef et al. 1997; Hodkinson et al. 2002b). In the scientific literature, this predominant genotype is often referred to by only its species name, but in other cases it has been given different genotypic identifiers. Given that *M.* ×*giganteus* is a species, of which there are an infinite number of potential genotypes, we propose that the predominant genotype currently grown in the West, and described by Hodkinson and Renvoize (2001) and Greef and Deuter (1993) be referred to as *M.* ×*giganteus* 'Aksel Olson' (see Darke 2007). 'Aksel Olson', and perhaps a few sports of this genotype, has been the foundation of nearly all efforts in Europe and the USA to establish a commercial *Miscanthus* biomass industry. Deuter and Abraham (1998) reported that another triploid genotype called *M.* ×*giganteus* 'Harvey' was imported from Japan to England in the 1980s. Such a narrow germplasm base for a crop is a great risk, as susceptibility to a single pest or pathogen could destroy all commercial plantings.

Given the confusion in the literature about the definition of *M.* ×*giganteus*, it is worth reviewing that the International Code of Botanical Nomenclature (McNeill et al. 2006) is clear that a nothospecies is defined by its parental species and not by its ploidy or genomic composition (see the *Oenothera* example in article H.4, Appendix 1). Backcrosses to either parental species are also included in the nothospecies definition (McNeill et al. 2006). Therefore, any interspecific combination of *M. sinensis* and *M. sacchariflorus*, regardless of whether the progeny are diploid, triploid, tetraploid or higher, are by definition correctly named *M.* ×*giganteus*. Thus, to the extent that the indigenous *Miscanthus* polyploids of Japan, Ogi, are known to be allopolyploids derived from *M. sinensis* and *M. sacchariflorus*, these too are most accurately described as *M.* ×*giganteus* (unless of course an earlier legitimate name is found in the literature). The name *M. ogiformis* has been used to describe hybrids of tetraploid Ogi and diploid *M. sinensis* (Hirayoshi et al. 1957; Honda 1939; Linde-Laursen 1993) but Hodkinson et al. (2002c) chose not to use this name in place of the newer *M.* ×*giganteus*. Honda (1939) provided a Latin description and a type specimen for *M. ogiformis* but he did not directly define the species as a hybrid of *M. sacchariflorus* and *M. sinensis*, though he indicated by its Japanese name, Ogi-susuki, that it was thought to be such a hybrid (Susuki is the common name of *M. sinensis* in Japanese). This leads us to conclude that *M.* ×*giganteus*, which is unambiguously defined in Latin as a hybrid of *M. sinensis* and *M. sacchariflorous* (Hodkinson and Renvoize 2001), is the appropriate name to use for such hybrids. Given this conceptual framework of *M.* ×*giganteus* as an allopolyploid of *M. sinensis* and *M. sacchariflorus*, it will be important for workers to know the ploidy level of the genotypes they study and communicate this to others.

The indigenous *M.* ×*giganteus* of Japan are typically late flowering, short-day plants, with stout rhizomes. They are typically tetraploid, though rarely triploid or pentaploid (Adati and Shiotani 1962; Hirayoshi et al. 1957). In the northern part of their range, they rarely, if ever, produce seed but instead propagate vegetatively by rhizomes and aboveground axillary buds. The genetic source of short-day flower

initiation in Japanese *M.* ×*giganteus* is unknown but an understanding of how this evolved would help guide plant breeders who wish to remake the cross to develop improved cultivars. One possibility is that a short-day *M. sinensis* from the tropics or subtropics of Southeast Asia combined with northern-adapted *M. sacchariflorus*, perhaps with the aid of a typhoon to bring seed of the *M. sinensis* parent northward. Another possibility is that *M. sacchariflorus* var. *lutarioriparius* combined with a northern-adapted day-neutral *M. sinensis*. In either scenario diploid progenies with some fertility would have likely been produced. Perhaps chromosome doubling improved chromosome pairing and fertility while establishing a high degree of genetic isolation from the parental species, and thereby providing a selective advantage for the tetraploid. Molecular marker and cytogenetic studies of germplasm from China, Korea and Japan will be needed to obtain a clearer understanding of how *M.* ×*giganteus* evolved.

2.6 *M. tinctorius*

M. tinctorius, is a tufted, shade-tolerant plant with narrow leaves that has a limited geographic distribution in central Honshu, Japan, primarily along the west coast between 1,000 and 1,500 m (Adati and Shiotani 1962; Hirayoshi et al. 1957; Ibaragi and Ohashi 2004; Lee 1964c; Matumura and Yukimura 1975). Its Japanese name, Kariyasu, means easy to cut, in reference to its traditional use for fodder (Hirayoshi et al. 1957; Matumura and Yukimura 1975). As its specific epithet indicates, it also has traditionally been used for making dyes, especially brilliant yellows (Darke 2007; Watanabe and Takahashi 2006). Perhaps the trait of potentially greatest value in *M. tinctorius* is its notably short callus hairs, which are only 1–2 mm long and usually half or less than the length of the spikelet (Clayton et al. 2010; Sun 2009). Callus hairs catch the wind and thereby disperse the seed. Loss of natural seed-dispersal mechanisms is a common feature of crop domestication and it would be desirable to limit seed dispersal of *Miscanthus* as we domesticate it to be a biomass crop.

2.7 *Saccharum*

As discussed in detail by Moore et al. (Chapter 3, this volume), modern sugarcane cultivars are complex hybrids of *S. officinarum*, *S. robustum* and *S. spontaneum*; they are highly polyploid and often aneuploid yet they are typically fertile (Burner 1997; Ramdoyal and Badaloo 2002). During the past 70 years, crosses between sugarcane cultivars and *Miscanthus* have been made many times, primarily to introgress genes for disease-resistance into sugarcane (Chen 1993; Chen and Lo 1989; Grassl 1974; Li et al. 1948, 1951, 1961; Xiao and Tai 1994). In addition, natural hybrids between sugarcane and *Miscanthus* have been reported in the South Pacific

(Grassl 1959). Unreduced gametes have been an important feature of *Saccharum*/*Miscanthus* crosses. Progeny of *Saccharum*/*Miscanthus* have been produced from gametes that were: $2n+n$, $n+n$, and $n+2n$ (Burner 1997; Chen et al. 2000; Li et al. 1948, 1951, 1953; Loh and Wu 1949). The type of progeny produced was influenced by parental genotype (Burner 1997; Li et al. 1948, 1953). In $2n+n$ progeny, characteristics of the *Saccharum* parent, such as thick culms, predominated (Li et al. 1948, 1953). In contrast, $n+n$ and $n+2n$ progeny had more *Miscanthus* traits, such as culm diameter that was intermediate but biased towards the thinner *Miscanthus* parent (Li et al. 1948, 1953). Progeny have also been produced from the reciprocal cross, *Miscanthus*/*Saccharum* (Jorge Da Silva, personal communication). These intergeneric progeny, which are sometimes referred to as miscanes, vary in fertility but have been used successfully in further crosses. Chen et al. (2000) reported using an F_1 miscane as a male parent to backcross disease-resistance genes into sugarcane. In subsequent generations, chromosomes of either parent can be lost. Grassl (1974) obtained hybrid progeny when he crossed *Miscanthus* as the female parent with *S. arundinaceum*, *S. ravennae*, or *S. procerum*. Given that sugarcane has been successfully crossed to *Sorghum* multiple times (de Wet et al. 1976; Gupta et al. 1978), it is not inconceivable that hybrids between *Miscanthus* and *Sorghum* might someday be made, either directly or in combination with *Saccharum*.

3 *Miscanthus* Improvement and Uses

3.1 Traditional Uses

In its native range, Miscanthus historically has been used for a wide variety of purposes including, long-lasting thatch, grazing, and fodder (Christophersen 1935; Iketani and Ida 2008; Matumura and Yukimura 1975; Stewart et al. 2009). In some cases plants were harvested from wild stands but in other cases people have managed landscapes by burning, harvesting or grazing to encourage growth of *Miscanthus*. Japanese grasslands dominated by *M. sinensis* have been managed by burning for several thousand years (Miyabuchi and Sugiyama 2006; Ogura et al. 2002), which was apparently originally due to the need to facilitate successful hunting of wild game by indigenous peoples. However, the need for managing the grasslands of Japan then shifted to supplying a source of long-lasting thatching material for houses and traditional religious structures (Koyama 1987; Iketani and Ida 2008; Yoshida et al. 2008). These grasslands were also harvested as fodder for livestock. At least in one *M. sinensis*-dominated grassland on Kyushu Island in southern Japan, grass is still harvested as feed for native Japanese breeds of cattle (Wagyu) (Nakaboh, personal communication). Interestingly, the primary use of *M. sinensis*-dominated grasslands in Japan is currently for tourism (Gartelmann 2001). In New Guinea and Taiwan, young *Miscanthus* shoots are sometimes used for human consumption (AVRDC 2003; French 2006). In Europe and the USA, *Miscanthus* has

Fig. 4.2 Ornamental cultivars of *Miscanthus sinensis* (*foreground*) and *M.* ×*giganteus* "UIUC" (*background*) in Champaign, IL

been grown, as an ornamental garden plant, from Victorian through to modern times (Darke 1994).

M. tinctorius, which has been traditionally used for making dyes, may be furthest along the path to domestication of the *Miscanthus* species, as evidenced by its very short callus hairs, which reduce its capacity to propagate, and by recent observations that its numbers in the wild are dwindling as it intercrosses with *M. oligostachyus* (Sun 2009; H. Watanabe, personal communication). Efforts to conserve *M. tinctorius* in its pure form should be undertaken.

3.2 Breeding Strategies for Bioenergy Feedstocks

During the past century there have been a few efforts to breed *Miscanthus* for a variety of purposes. In China, *M. sacchariflorus* var. *lutarioriparius* was bred as a feedstock for paper production. In Japan, a breeding program was initiated during the 1950s to develop *Miscanthus*, especially *M. sinensis* and *M. sinensis*/*M. sacchariflorus* hybrids, as a forage crop (Adati 1958b; Hirayoshi et al. 1960). Matumura et al. (1985, 1987) continued the work on triploid and tetraploid *M. sinensis*/*M. sacchariflorus* hybrids started by Hirayoshi et al. (1960). In Europe, a considerable breeding effort, especially by nurseryman Ernst Pagels (Darke 1994, 2007), to develop ornamental *Miscanthus* cultivars has made *Miscanthus* a common garden plant throughout Europe and North America. Many of the newer ornamental

cultivars have been bred for short height and early flowering, the exact opposite of what is needed for a biomass cultivar (Fig. 4.2).

Current interest in *Miscanthus* biomass is focused primarily on sustainable production of renewable energy (Clifton-Brown et al. 2008). This goal will require efforts to domesticate *Miscanthus*. Traits that prevent the crop from spreading beyond cultivated fields, such as nonshattering, absence of callus hairs, or sterility will be selected. If seed-propagated cultivars are bred, these will be selected for relatively large seed, uniform timing of flowering, and high seed yield (Matumura and Yukimura 1975). Clonally propagated cultivars will be selected for a tufted habit or at most, a moderately spreading habit, and for ease of vegetative propagation. Selection for adaptation to target biomass-production environments will be a major theme. Adaptive traits will likely include tolerance to cold, drought, waterlogged or saline soils, as well as resistance to pests and diseases. Yield and yield components, such as height, tiller density, culm diameter and flowering time, will be of central importance. Quality traits such as low ash content, reduction of minerals such as Cl and K that can be corrosive to combustion equipment, shedding of mineral-rich leaves prior to harvest, dormancy-associated transport of nutrients belowground, ratios of lignin, cellulose, and hemicellulose, and low moisture content at harvest will also be of value. Timing of dormancy and nutrient recycling will be important for sustainability (Himken et al. 1997). Experience with *Miscanthus* and other crops, including other Saccharinae, indicates that manipulation of flowering time and photoperiod response will have a major impact on domestication, adaptation, and yield (Clifton-Brown et al. 2008; Harlan 1992).

As a relatively unimproved biomass crop, the introduction of genetic variation has the potential to produce substantial increases in *Miscanthus* yield and quality, and to provide sources of resistance to pests and diseases that may arise in the future (Clifton-Brown et al. 2001a). Unfortunately there is almost no genetic variation among the few (three or four) accessions of triploid *M.* ×*giganteus* currently available in Europe and the USA (Greef et al. 1997). However, significant genetic variability has been found among the parental species, *M. sinensis* and *M. sacchariflorus* (Jorgensen and Muhs 2001).

So what approaches are available to plant scientists to tap the genetic diversity of *Miscanthus* germplasm for crop improvement and avoid the vulnerability and constrained adaptability of a production system based on a single clonally propagated genotype? Potential strategies for *Miscanthus* crop improvement include the following: (1) resynthesis of new triploid *M.* ×*giganteus* genotypes by conventional hybridization, (2) intra- and interspecific hybridization and progeny selection among diploid accessions of *M. sinensis*, *M. sacchariflorus*, and *M. floridulus*, (3) manipulation of ploidy levels to circumvent reproductive barriers between species and make new genomic combinations, and (4) genetic modification.

3.3 *M. ×giganteus: Lessons Learned*

The great potential value of crosses between *M. sinensis* and *M. sacchariflorus* for breeding high-biomass *Miscanthus* has long been known. Recent interest in *Miscanthus* for bioenergy has led to increased work in this area during the last decade. Hirayoshi et al. (1960) crossed a diploid *M. sinensis* var. *condensatus* as a female parent with a natural tetraploid *M. ×giganteus* (Japanese Ogi) male and obtained one triploid and one tetraploid progeny, concluding that the tetraploid progeny was the product of an unreduced female gamete. Notably, Hirayoshi et al. (1960) observed transgressive segregation for biomass traits in both the triploid and the tetraploid progeny. As might be expected, the triploid progeny was sterile and looked more like the tetraploid *M. ×giganteus* (Japanese Ogi) parent but the tetraploid progeny was fertile and looked more like the *M. sinensis* parent (Hirayoshi et al. 1960). Adati (1958a) noted that naturally occurring pentaploid *M. ×giganteus* (Ogi) were highly vigorous. Matumura et al. (1985, 1986, 1987) conducted a series of detailed field studies on the growth and development of the triploid and tetraploid progenies of Hirayoshi et al. (1960) and found that the tetraploid progeny out-yielded its parents and triploid sibling by more than 2:1 on a per area basis. The higher yield of the tetraploid progeny was partially due to optimal tiller density and leaf area index. Matumura et al. (1985, 1987) indicated that the triploid and tetraploid progenies could be used as new forage cultivars and, moreover, they suggested a breeding program based on their work to develop further improved forage cultivars. These reports from Japan, taken together with the success of a triploid genotype of *M. ×giganteus* in Europe, suggest that triploid, tetraploid, and pentaploid *M. ×giganteus* will be promising avenues of research to develop improved biomass cultivars.

Recent research efforts in Europe have also focused on crosses between *M. sinensis* and *M. sacchariflorus*. Clifton-Brown et al. (2008) reported that a breeding program at the Aberystwyth University in the UK and Plant Research International in the Netherlands was initiated in 2004 and is focusing on improvement of *M. sinensis* and using their *M. sinensis* selections to develop improved versions of *M. ×giganteus*. A German breeding program led by Martin Deuter at Tinplant Biotechnik und Pflanzenvermehrung GmbH was established in 1992 and released two cultivars in 2006 *M. ×giganteus* 'Amuri' and 'Nagara' (http://www.tinplant-gmbh.de/). As suggested by their cultivar names, 'Amuri' was derived from a cross between North Asian *M. sacchariflorus* and *M. sinensis*, while 'Nagara' was derived from a Japanese *M. ×giganteus* (described as *M. sacchariflorus*) crossed with *M. sinensis* (http://www.tinplant-gmbh.de/; http://renewable-energy.illinois.edu/docs/symposium/PDF/Zhang_Talk.pdf). 'Amuri' and 'Nagara' are currently being marketed in Canada (www.newenergyfarms.com). Promising hexaploid *M. ×giganteus* individuals have recently been produced at the University of Illinois by chromosome-doubling a triploid genotype (referred to as 'UIUC') that has been widely studied in the USA (Yu et al. 2009).

3.4 Resynthesis of New Triploid *M.* ×*giganteus* Genotypes

The outcome of future field trials of *M.* ×*giganteus* 'Nagara' notwithstanding, existing *Miscanthus* crop improvement programs in Europe and the USA have yet to develop germplasm superior to the single genotype currently in production in Europe. This is due in part to limited availability of parental germplasm in the West, particularly accessions of tetraploid *M.* ×*giganteus* from Japan (Stewart et al. 2009), difficulties in generating triploid progeny from crosses between a short-day tetraploid and day-neutral diploid parents (Clifton-Brown, personal communication), and the inability to generate progeny from the triploid genome of *M.* ×*giganteus.* These limitations can and should be addressed with the development of extensive public germplasm collections, and by the use of greenhouse and growth room facilities that control day-length. Naturally occurring triploid *M.* ×*giganteus* may also be obtained from areas in Japan where sympatric populations of *M. sinensis* and tetraploid *M.* ×*giganteus* exist (Hirayoshi et al. 1957; Nishiwaki et al. 2011. Though the current *M.* ×*giganteus* genotype of commerce in Europe and the USA is an excellent cultivar, it is improbable that this chance find is the best that can be produced. By choosing outstanding tetraploid and diploid parents, superior triploid progeny should be obtained. Moreover, such crosses may also yield useful tetraploid progenies via unreduced gametes, as the work of Hirayoshi et al. (1960) and Matumura et al. (1985, 1987) has shown.

3.5 Intra- and Interspecific Hybridization Among Diploids

Crosses among *Miscanthus* diploids could be used to develop biomass cultivars directly or as a stepping stone to developing superior triploids by selecting for improved parents. While triploid *M.* ×*giganteus* poses less invasive potential as a bioenergy crop due to sterility, it is not amenable to conventional breeding. In addition, establishment of triploid *M.* ×*giganteus* production fields requires vegetative propagation via rhizomes, divisions or tissue culture, which would be expensive relative to seed propagation if the latter option were available. For rhizome harvest, cleaning, separation and replanting, it has been estimated that one hectare of mature *M.* ×*giganteus* (3 or more years old) will provide sufficient rhizomes for the planting of approximately only ten hectares of new production (T. Voigt, personal communication). This is a significant constraint in the development of a bioenergy cropping system that will need to fulfill large production requirements. In contrast, seed-based propagation of *Miscanthus* should be much cheaper, scalable to meet production needs, and utilize sexual hybridization and selection to generate diverse and improved germplasm for commercial production.

Since all known accessions of *Miscanthus* are self-incompatible, individual plants tend to be highly heterozygous and seed generated from hybridizations can display substantial segregation. Controlled pollinations with *M. sinensis* can be

conducted either by collecting and applying pollen from a male parent(s) followed by bagging of flowering inflorescences of a female(s) parent to prevent contamination, or by isolation. Isolating together two genotypes that flower concurrently can produce seed from known male and female parents. This provides the opportunity for plant breeders to use mass selection and full or half-sib selection for crop improvement. One breeding strategy currently in use by forage breeders that could be applied to *Miscanthus* crop improvement is the development of synthetic cultivars that consist of a seed mixture generated from superior performing parental genotypes by open pollination. Synthetic hybrids, produced by using just two clonally propagated noninbred self-incompatible parents for seed production has been proposed by Taliaferro et al. (1999) for switchgrass and could be useful for *Miscanthus* too.

A major limitation in *Miscanthus* breeding is the need to grow out segregating populations for at least 2 or 3 years to evaluate potential individual genotypes for use in a synthetic for commercial production. It was observed that for autumn biomass of 15 *Miscanthus* genotypes (*M. sinensis*, *M. sacchariflorus,* and *M.* ×*giganteus*) measured during the first 3 years after establishment there was a correlation of 0.81 between second and third year yields compared to a correlation of only 0.56 between first and third year yields (Clifton-Brown et al. 2001b). This need for mature stand evaluation will significantly extend the time needed to develop commercial cultivars. Availability of germplasm for breeding in the USA is also a limiting factor at this time. Though some limited clonal germplasm of *M. sinensis* is available for purchase from nurseries, availability of germplasm of the other *Miscanthus* species is currently lacking or severely constrained.

Interspecific crosses among diploids may be a useful strategy for capturing heterosis and transgressive segregation in a seeded product. Crosses between diploid *M. sinensis* and *M. sacchariflorus* have produced selections with adaptation to cold northern environments, high yield and early maturity (energy.illinois.edu/docs/symposium/PDF/Zhang_Talk.pdf). Adati and Shiotani (1962) reported on a cross between *M. sinensis* and *M. floridulus* that yielded hybrid progeny that showed regular meiotic pairing of 19 bivalents in the first metaphase. This information suggests that hybridization between diploid germplasm of each of the species could provide access to the extensive genetic diversity found within the genus *Miscanthus.*

M. sacchariflorus var. *lutarioriparius*, given its current commercial use in China for paper production (Xi 2000), will be of interest for breeding bioenergy feedstock cultivars for China, the USA and Europe (Xi 2003). However, in contrast to typical forms of *M. sacchariflorus*, such as the cultivar 'Robustus', researchers in Europe and the USA have had little experience with *M. sacchariflorus* var. *lutarioriparius*. Given its fertility and spreading habit, *M. sacchariflorus* var. *lutarioriparius* should not be used directly as a biomass crop in the southern USA, where it has the potential to escape cultivations and disturb natural riparian environments. Though F_1 *M. sinensis*/*M. sacchariflorus* var. *lutarioriparius* hybrids are typically tufted in habit, use of these directly for crop development or further breeding should proceed with caution. Our current understanding of rhizome

genetics in the *Poaceae* (Hu et al. 2003; Paterson et al. 1995) suggests that two or more backcrosses to the tufted *M. sinensis* would produce progeny with a very low probability of having a spreading habit. Nevertheless, sterility would be a desirable trait for future cultivars derived from *M. sacchariflorus* var. *lutarioriparius* that are to be grown outside of its native range.

3.6 Manipulation of Ploidy Levels

Chromosome doubling is a viable approach to generate polyploid plants that can have the advantage of larger cell size and higher vegetative biomass yield (Acquaah 2007). New polyploids commonly exhibit gigantism with sturdier foliage, thicker stems, and enlarged reproductive structures (Ramsey and Schemske 1998), although reduced tillering has been observed in tetraploid rye (Muntzing 1951), tetraploid rice (Tu et al. 2007) and allohexaploid pearl millet × napiergrass hybrids (Gonzalez and Hanna 1984). Artificial chromosome doubling has been used to create fertile amphidiploids or hexaploids from sterile hybrids of related plant species (Thomas 1993; Nimura et al. 2006). This approach has been successfully used to mine germplasm of incompatible diploid or polyploid species related to cultivated potato (Carputo and Barone 2005), pearl millet (Hanna 1990), sorghum (Luo et al. 1992) and other crops. Chromosome doubling of related diploid species has also been repeatedly used to overcome incompatibility to generate fertile interspecific tetraploid hybrids (Ramsey and Schemske 1998). Hybridization of induced autotetraploids of *indica* and *japonica* subspecies of rice displayed enhanced heterosis and fertility making them amenable to conventional sexual hybridization for crop improvement (Tu et al. 2007). Chromosome doubling of a sterile triploid interspecific rose hybrid generated a hexaploid plant with restored pollen viability. This allohexaploid was also able to undergo self-pollination and set viable seed (Kermani et al. 2003). These studies suggest that ploidy manipulation via chromosome doubling could access genetic variability for the improvement of *Miscanthus* as a biofuel/bioenergy crop.

Chromosome doubling commonly involves treatment of tissue-cultured callus or shoots with antimitotic agents like amiprophos-methyl, colchicine, oryzalin, or trifluralin, from which polyploid plantlets are then regenerated. Kim et al. (2010) has developed a successful *Miscanthus* regeneration system patterned after those used before with *Miscanthus* (Holme and Petersen 1996; Holme et al. 1997) and regenerated hundreds of *M.* ×*giganteus* plants from callus of immature inflorescence tissue. Plants of both diploid *M. sinensis* and *M. sacchariflorus* have also been regenerated from callus from hypocotyls of germinating seeds and immature inflorescence tissue.

Using this regeneration system approximately 35 % of *M.* ×*giganteus* calli treated with the chromosome-doubling agents, colchicine and oryzalin, had nuclear DNA content twice that observed in untreated triploid controls (Yu et al. 2009).

Fig. 4.3 Triploid *M.* × *giganteus* 'UIUC' (*left*), and an induced hexaploid *M.* ×*giganteus* (*right*)

Ten *M.* ×*giganteus* plants were recovered that when measured for DNA content by flow cytometry were hexaploids via chromosome doubling. Preliminary phenotypic comparison of control (triploid) and hexaploid plants suggests that the hexaploids have slightly broader stems and show slightly greater growth rates, but may display reduced tillering (Fig. 4.3). Field trials of the hexaploids are currently underway. In a greenhouse, the hexaploid plants have initiated flowering, generating pollen, a proportion of which after treatment with fluorescein diacetate and assayed for UV fluorescence (Ueda 1994) has proven to be viable (W.B. Chae, personal communication; Fig. 4.4). This same protocol has also been used to regenerate tetraploid *M. sinensis* and *M. sacchariflorus* plant accessions from diploid tissues (W.B. Chae, personal communication).

The potential to generate viable seed from polyploid hybridizations between and among accessions of tetraploid *M. sinensis* and *M. sacchariflorus* and hexaploid *M.* ×*giganteus* may help to unlock the genetic variation available within *Miscanthus* to drive a conventional breeding program for the development of superior feedstock

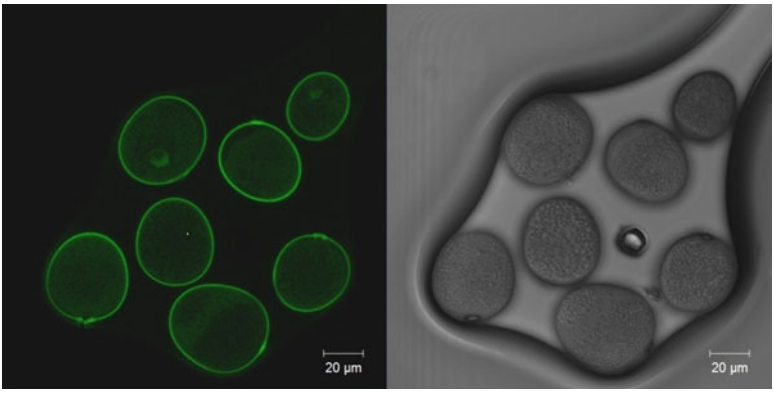

Fig. 4.4 Fluorescein diacetate staining of hexaploid *M.* ×*giganteus* pollen indicating viability (*left*: yellow-green fluorescing pollen grains under ultraviolet light; *right*: the same pollen grains viewed without ultraviolet light)

germplasm adapted to a range of environmental conditions. New allopolyploids can be produced by chromosome-doubling of diploid parents or their interspecific progenies. Newly developed allopolyploids derived from *M. sinensis* and *M. sacchariflorus* could be used to broaden the genetic base of natural *M.* ×*giganteus*, thereby allowing for further exploitation of heterosis. Crosses between tetraploids and diploids or between hexaploids and tetraploids could generate triploid and pentaploid seed that could result in sterile plants for biomass production systems that would reduce concerns of invasiveness.

3.7 Genetic Modification

Though conventional breeding may be able to manipulate a number of traits for *Miscanthus* species, there is very little presently known about what trait variation is available in the germplasm and the breeding systems are just being developed. *M.* ×*giganteus* is sterile, so breeding cannot be readily accomplished with these clones and little variability is also apparent. Thus being able to insert genes appears to have some real importance. Most plant transformation utilizes tissue cultures and methods have been published for culture initiation, maintenance and plant regeneration as stated above (Holme and Petersen 1996; Kim et al. 2010). Usually the gene of interest and selectable marker gene are inserted into cells by particle bombardment or A*grobacterium tumefaciens* cocultivation. The transformed cells are selected using a selective agent that kills untransformed cells, but not those expressing the selectable marker gene, such as antibiotic resistance. Plants are then regenerated. Transformation of *Miscanthus* by *Agrobacterium tumefaciens*, and by particle bombardment has been reported, with successful transformation obtained for *M. sacchariflorus* and *M. sinensis* (Engler and Chen 2009; Wang et al. 2011; Yi et al. 2001).

Efforts are currently underway at the University of Georgia and the University of Illinois is to generate resources that will enable genomics-directed improvement of *Miscanthus* germplasm. The genome of *M.* ×*giganteus* is very large, estimated to be approximately 7.0 Gbp by flow cytometry (Rayburn et al. 2009). Using new generation genomic tools 1× skim sequencing of *M.* ×*giganteus* DNA revealed that much of the genome consists of major repeated sequences with only 2.4 % or about 165 Mbp as "genespace" and the recently sequenced *Sorghum bicolor* is a useful reference genome (Swaminathan et al. 2010). Deep sequencing of the *M. sinensis*, *M. sacchariflorus*, and *M.* ×*giganteus* transcriptome found that contigs matched approximately 29,000 of the estimated 36,000 *Sorghum* genes.

This sequence information has been used to generate SNPs (single nucleotide polymorphisms), SSRs (single sequence repeats) and other PCR-based markers that have been made available to the public for linkage studies, association mapping, and marker-assisted breeding. The identification of the most informative SNPs across the transcriptome of these species can be used to generate an Illumina GoldenGate SNP array to apply to *M. sinensis* and *M. sinensis*/*M. sacchariflorus* hybrid segregating populations. This can provide information to create genetic linkage maps for these species and be used to identify QTL and genes associated with desired phenotypes.

3.8 Energycanes and Miscanes

Past work on Miscanes focused primarily on the introgression of genes for disease-resistance from *Miscanthus* to sugarcane. Such backcross programs faced the difficult task of maintaining the resistance while recovering sugar production and other agronomic traits into a sugarcane genetic background. Future work to introgress cold tolerance from *Miscanthus* to low-sugar but high-biomass energy cane cultivars is expected to be less challenging than the earlier work that needed to recover genotypes with high sugar and low fiber. Miscanes also have the potential to become an important biomass crop in warm-temperate and subtropical regions by combining the best attributes of both parents. Sugarcane parents have genes for high biomass yield-potential, which is achieved through strong, thick culms, and great height mediated by late flowering, which is facilitated by a strong requirement for short and shortening days to initiate flowering (Glyn 2004). *Miscanthus* has genes for high yield potential, realized primarily via high culm density, which is complementary to sugarcane. Most critically, *Miscanthus* brings genes for adaptation to temperate environments, especially cold-tolerance. In addition, the dormancy and dry-down traits of *Miscanthus* could be especially valuable where the end-use is combustion. Burner et al. (2009) observed better cold tolerance and biomass yield in Arkansas for two miscane genotypes, in comparison to progenies of *S. spontaneum*, which is a traditional source of cold tolerance genes for breeding sugarcane. Extractable energy content ha^{-1} $year^{-1}$, obtained primarily from cellulose and lignin rather than sugar, is likely to be the most important criteria for future biomass cultivars of miscanes.

4 Conclusions

The substantial genetic diversity already observed within *Miscanthus* represents opportunities to develop it into an important biomass crop for meeting our renewable energy needs. A greater understanding of *Miscanthus* trait diversity, population structures, and evolution, especially in the context of other Saccharinae, is needed. To obtain a full and accurate understanding of *Miscanthus* evolution, future research must combine information from different disciplines, including population genetics, cytogenetics, molecular genetics, and genomics. To facilitate basic research and applied breeding, the research community needs to quickly establish and make publicly available *Miscanthus* germplasm collections that broadly represent the diversity of the genus.

References

Acquaah G (2007) Polyploidy in plant breeding. In: Chap 13, Principal of plant genetics and breeding. Blackwell Publishing, Malden, MA, pp 214–230

Adati S (1958a) Cytogenetics of Japanese wild forage *Miscanthus* species. In: Proceedings of the X International Congress of Genetics, McGill University, Montreal, Canada, 20–27 August

Adati S (1958b) Studies on the genus *Miscanthus* with special reference to the Japanese species for breeding purpose as fodder crops. Bull Fac Agric Mie Univ 12:1–112

Adati S, Shiotani I (1962) The cytotaxonomy of the genus *Miscanthus* and its phylogenic status. Bull Fac Agric Mie Univ 25:1–24

Amalraj VA, Balasundaram N (2006) On the taxonomy of the members of 'Saccharum complex'. Genet Resour Crop Evol 53:35–41

AVRDC (2003) Program III. Collaboration in research and germplasm management. In: Kalb T (ed) AVRDC Report 2002, Taiwan

Beale CV, Long SP (1997) Seasonal dynamics of nutrient accumulation and partitioning in the C-4 grass *Miscanthus* × *giganteus* and *Spartina cynosuroides*. Biomass Bioenergy 12:419–428

Beale CV, Bint DA, Long SP (1996) Leaf photosynthesis in the C4-grass Miscanthus × giganteus, growing in the cool temperate climate of southern England. J Exp Bot 47:267–273

Beale CV, Morison JI, Long SP (1999) Water use efficiencies of c4 perennial grasses in a temperate climate. Agric Forest Meteorol 96:103–115

Burner DM (1997) Chromosome transmission and meiotic behavior in various sugarcane crosses. J Am Soc Sugar Cane Technol 17:38–50

Burner DM, Tew TL, Harvey JJ, Belesky DP (2009) Dry matter partitioning and quality of *Miscanthus*, *Panicum*, and *Saccharum* genotypes in Arkansas, USA. Biomass Bioenergy 33:610–619

Carputo D, Barone A (2005) Ploidy level manipulations in potato through sexual hybridization. Ann Appl Biol 146:71–79

Chen YH (1993) Genetics and breeding studies on *Saccharum-Miscanthus* nobilization. Dissertation, National Taiwan University

Chen YH, Lo CC (1989) Disease resistance and sugar content in *Saccharum-Miscanthus* hybrids. Taiwan Sugar 36:9–12

Chen SL, Renvoize SA (2005) A new species and a new combination of Miscanthus (Poaceae) from China. Kew Bull 60:605–607

Chen YH, Chen C, Lo CC (2000) Extraordinary phenomenon of cell division in *Saccharum Miscanthus* and their nobilized progenies. Rep Taiwan Sugar Res Inst 170:27–44

Chiang YC, Chou CH, Huang S, Chiang TY (2003a) Possible consequences of fungal contamination on the RAPD fingerprinting in *Miscanthus* (Poaceae). Aust J Bot 51:197–201
Chiang YUC, Schaal BA, Chou CH, Huang S, Chiang TY (2003b) Contrasting selection modes at the ADH1 locus in outcrossing *Miscanthus sinensis* vs. inbreeding *Miscanthus condensatus* (Poaceae). Am J Bot 90:561–570
Chou CH, Chiang YC, Chiang TY (2000) Genetic variability and phytogeography of *Miscanthus sinensis* var. *condensatus* an apomictic grass based on RAPD fingerprints. Can J Bot 78:1262–1268
Chou CH, Chiang TY, Chiang YC (2001) Towards an integrative biology research: a case study on adaptive and evolutionary trends of *Miscanthus* populations in Taiwan. Weed Biol Manage 1:81–88
Christian DG, Haase E (2001) Agronomy of miscanthus. In: Jones MB, Walsh M (eds) Miscanthus for energy and fibre. James & James, London, pp 21–45
Christophersen E (1935) Flowering plants of Samoa. Bernice P. Bishop Museum Bull 128, Honolulu
Clayton WD, Harman KT, Williamson H (2010) GrassBase - the online world grass flora. http://www.kew.org/data/grasses-db.html
Clifton-Brown JC, Jones MB (2001) Yield performance of M. ×giganteus during a 10 year field trial in Ireland. Aspects Appl Biol 65:153–160
Clifton-Brown JC, Long SP, Jorgensen U (2001a) *Miscanthus* productivity. In: Jones MB, Walsh M (eds) Miscanthus for energy and fibre. James & James, London, pp 46–67
Clifton-Brown JC, Lewandowski I, Andersson B, Basch G, Christian DG, Kjeldsen JB, Jørgensen U, Mortensen JV, Riche AB, Schwarz KU, Tayebi K, Teixeira F (2001b) Performance of 15 *Miscanthus* genotypes at five sites in Europe. Agron J 93:1013–1019
Clifton-Brown JC, Stampfl PF, Jones MB (2004) *Miscanthus* biomass production for energy in Europe and its potential contribution to decreasing fossil fuel carbon emissions. Glob Chang Biol 10:509–518
Clifton-Brown J, Chiang Y-C, Hodkinson TR (2008) Miscanthus: genetic resources and breeding potential to enhance bioenergy production. In: Vermerris W (ed) Genetic improvement of bioenergy crops. Springer, New York, pp 273–294
Darke R (1994) A century of grasses. Arnoldia 54:3–11
Darke R (2007) The encyclopedia of grasses for livable landscapes. Timber, Portland
de Wet JMJ, Gupta SC, Harlan JR, Grassl CO (1976) Cytogenetics of introgression from *Saccharum* into *Sorghum*. Crop Sci 16:568–572
Deuter M, Abraham J (1998) Genetic resources of *Miscanthus* and their use in breeding. In: Biomass for energy and industry proceedings of the international conference. 10th European conference and technology exhibition, Wurzburg, Germany, 8–11 June 1998
Dohleman FG, Long SP (2009) More productive that maize in the midwest: how does *Miscanthus* do it? Plant Physiol 150:2104–2115
Engler D, Chen J (2009) Transformation and engineered trait modification in miscanthus species. World Intellectual Property Organization
Florence J, Lorence DH (1997) Introduction to the flora and vegetation of the Marquesas Islands. Allertonia 7:226–237
French BR (2006) Food composition tables for food plants in Papua New Guinea. Tasmania
Gartelmann S (2001) Where there's a spark, there's green tourism. The Japan Times Online URL. http://search.japantimes.co.jp/cgi-bin/fv20010321a1.html
Glyn JL (2004) An introduction to sugarcane. In: Glyn J (ed) Sugarcane, 2nd edn. Blackwell, Ames, pp 1–19
Gonzalez B, Hanna W (1984) Morphological and fertility responses in isogenic triploid and hexaploid pearl millet × napiergrass hybrids. J Hered 75:317–318
Grassl CO (1959) Introgression between *Saccharum* and *Miscanthus* in New Guinea and the Pacific area. In: Proceedings of the IX international botanical congress, Montreal, Canada, 19–29 August

Grassl CO (1974) The origin of sugarcane. Sugarcane Breed Newsl 34:10–18
Greef JM, Deuter M (1993) Syntaxonomy of *Miscanthus* × *giganteus* Greef et Deu. Angew Bot 67:87–90
Greef JM, Deuter M, Jung C, Schondelmaier J (1997) Genetic diverstiy of European *Miscanthus* species revealed by AFLP fingerprinting. Genet Resour Plant Evolut 44:185–195
Gupta SC, Harlan JR, de Wet MJ (1978) Cytology and morphology of a tetraploid zorghum population recovered from a *Saccharum* × *Sorghum* hybrid. Crop Sci 18:879–883
Haberle S (2007) Prehistoric human impact on rainforest biodiversity in highland New Guinea. Philos Trans R Soc B 362:219–228
Hanna WW (1990) Transfer of germplasm from the secondary to the primary pool in *Pennisetum*. Theor Appl Genet 80:200–204
Harlan JR (1992) Crops and Man. Am Soc Agron, Madison, WI
Heaton E, Voigt T, Long SP (2004) A quantitative review comparing the yields of two candidate C_4 perennial biomass crops in relation to nitrogen temperature and water. Biomass Bioenergy 27:21–30
Heaton E, Dohleman FG, Long SP (2008) Meeting US biofuel goals with less land: the potential of Miscanthus. SP Global Change Biol 14:1–15
Hictchcock AS (1971) Manual of grasses of the United States, vol 2, 2nd edn. Dover, New York
Himken M, Lammel J, Neukirchen D, Czypionka-Krause U, Olfs HW (1997) Cultivation of *Miscanthus* under West European conditions: seasonal changes in dry matter production, nutrient uptake and remobilization. Plant Soil 189:117–126
Hirayoshi I, Nishikawa K, Kato R (1955) Cytogenetical studies on forage plants. (IV) Self-incompatibility in *Miscanthus*. Jpn J Breed 5:167–170
Hirayoshi I, Nishikawa K, Kubono M, Murase T (1957) Cyto-genetical studies on forage plants (VI) On the chromosome number of Ogi (M*iscanthus sacchariflorus*). Res Bull Fac Agric Gifu Univ 8:8–13
Hirayoshi I, Nishikawa K, Kubono M, Sakaida T (1959) Cyto-genetical studies on forage plants (VII) Chromosome conjugation and fertility of *Miscanthus* hybrids including M. *sinensis M. sinensis* var. *condensatus* and *M. tinctorius*. Res Bull Fac Agric Gifu Univ 11:86–91
Hirayoshi I, Nishikawa K, Hakura A (1960) Cyto-genetical studies on forage plants (VIII) 3x- and 4x-hybrid arisen from the cross *Miscanthus sinensis* var. *condensatus* × *Miscanthus sacchariflorus*. Res Bull Fac Agric Gifu Univ 12:82–88
Hodkinson TR, Renvoize S (2001) Nomenclature of *Miscanthus* × *giganteus* (*Poaceae*). Kew Bull 56:759–760
Hodkinson TR, Chase MW, Lledó MD, Salamin N, Renvoize SA (2002a) Phylogenetics of *Miscanthus Saccharum* and related genera (Saccharinae Andropogoneae Poaceae) based on DNA sequences from ITS nuclear ribosomal DNA and plastid *trnL* intron and *trnL-F* intergenic spacers. J Plant Res 115:381–392
Hodkinson TR, Chase MW, Renvoize SA (2002b) Characterization of a genetic resource collection for *Miscanthus* (Saccharinae Andropogoneae Poaceae) using AFLP and ISSR PCR. Ann Bot 89:627–636
Hodkinson TR, Renvoize SA Chase MW (1997) Systematics of *Miscanthus*. Aspects of Applied Biology 49:189–198
Holme IB, Petersen KK (1996) Callus induction and plant regeneration from different explant types of *Miscanthus* × *ogiformis* Honda 'Giganteus'. Plant Cell Tissue Organ Cult 45:43–52
Holme IB, Petersen KK (1996) Callus induction and plant regeneration from different explant types of *Miscanthus* × *ogiformis* Honda 'Giganteus'. Plant Cell Tissue Organ Cult 45:43–52
Holme IB, Krogstrup P, Hansen J (1997) Embryogenic callus formation, growth and regeneration in callus and suspension cultures of *Miscanthus* × *ogiformis* Honda 'Giganteus' as affected by proline. Plant Cell Tissue Organ Cult 50:203–210
Honda M (1939) Nuntia ad Floram Japoniae. XXXVIII. Bot Mag Tokyo 53:144
Hu FY, Tao DY, Sacks E, Fu BY, Xu P, Li J, Yang Y, McNally K, Khush GS, Paterson AH, Li Z-K (2003) Convergent evolution of perenniality in rice and sorghum. Proc Natl Acad Sci U S A 100:4050–4054

Ibaragi Y, Ohashi H (2004) A taxonomic study of *Miscanthus* section *kariyasua* (*Graminae*). J Jpn Bot 79:4–22

Iketani Y, Ida H (2008) Flora of the grassland producing roof material in northern Nagano Prefecture, central Japan. Bull Inst Nat Educ Shiga Heights Shinshu Univ 45:1–6

Inthakoun L, Delang CO (2008) Lao Flora A checklist of plants found in Lao PDR with scientific and vernacular names. Lulu, Morrisville, NC

Iwanami Y (1969) Temperatures during *Miscanthus* type grassland fires and their effect on the regeneration of *Miscanthus sinensis*. Rep Inst Agric Res Tohoku Univ 20:47–88

Jensen EF (2009) Flowering time diversity in *Miscanthus*: a tool for the optimisation of biomass. Comp Biochem Physiol A: Mol Integr Physiol 153(2 (Suppl 1)):S197

Jorgensen U, Muhs H-J (2001) *Miscanthus* breeding and improvement. In: Jones MB, Walsh M (eds) Miscanthus for energy and fibre. James & James, London, pp 68–85

Kermani MJ, Sarasan V, Roberts AV, Yokoya K, Wentworth J, Sieber VK (2003) Oryzalin-induced chromosome doubling in *Rosa* and its effect on plant morphology and pollen viability. Theor Appl Genet 107:1195–1200

Kim C, Zhang D, Auckland SA, Rainville LK, Jakob K, Kronmiller B, Sacks EJ, Deuter M, Paterson AH (2012) SSR-based genetic maps of *Miscanthus sinensis* and *M. sacchariflorus*, and their comparison to sorghum. Theor Appl Genet 124:1325–1338

Kim HS, Zhang G, Juvik JA, Widholm JM (2010) Miscanthus × giganteus plant regeneration: Effect of callus types, ages and culture methods on regeneration competence. Global Change Biol Bioenergy 2:192–200

Koyama T (1987) Grasses of Japan and its neighboring regions: an identification manual. Kodansha Ltd, Tokyo

Lafferty J, Lelley T (1994) Cytogenetic studies of different *Miscanthus* species with potential for agricultural use. Plant Breed 113:246–249

Lee YN (1964a) Taxonomic studies on the genus *Miscanthus*: relationships among the section subsection and species part 1. J Jpn Bot 39:196–205

Lee YN (1964b) Taxonomic studies on the genus *Miscanthus*: relationships among the section subsection and species part 2 enumeration of species and varieties. J Jpn Bot 39:257–265

Lee YN (1964c) Taxonomic studies on the genus *Miscanthus*: relationships among the section subsection and species part 3 enumeration of species and varieties. J Jpn Bot 39:289–298

Lee YN (1993) Manual of the Korean grasses. Ewha Womans University Press, Seoul

Lewandowski I, Clifton-Brown JC, Scurlock JMO, Huisman W (2000) *Miscanthus*: European experience with a novel energy crop. Biomass Bioenergy 19:209–227

Li HW, Loh CS, Lee CL (1948) Cytological studies on sugarcane and its relatives I. Hybrids between *Saccharum officinarum Miscanthus japonicus* and *Saccharum spontaneum*. Bot Bull Acad Sin 2:147–160

Li HW, Ma TH, Shang KC (1951) Cytological studies of sugarcane and its relatives IX. Further studies of hybrids of intergeneric and interspecific crosses. Rep Taiwan Sugar Exp Stat 7:1–24

Li HW, Ma TH, Shang KC (1953) Cytological studies of sugarcane and its relatives X. Exclusive "patroclinous" type in the F_1 of sugarcane variety and *Miscanthus japonicus* Anders. Rep Taiwan Sugar Exp Stat 10:1–6

Li HW, Weng TH, Shang KC, Yang PC (1961) Cytological studies of sugarcane and its relatives: XVII. Trigeneric hybrids of *Saccharum officinarum* L. *Sclerostachya fusca* A. Camus and *Miscanthus japonicus* Anderss. Bot Bull Acad Sin 2:1–9

Linde-Laursen IB (1993) Cytogenetic analysis of *Miscanthus* 'Giganteus', an interspecifichybrid. Hereditas 119:297–300

Liu J, Yu X (2004) The exploitation and utilization of *Triarrhena lutarioriparia* resources. J Zhongkai Agrotechn Coll 7:63–67

Lledó MD, Renvoize SA, Chase MW (2001) *Miscanthus sinensis* and *Miscanthus sacchariflorus*: a confusing pair of species. Aspects Appl Biol 65:249–254

Loh CS, Wu TH (1949) A note on the trihybrids of (*Saccharum officinarum* × *S. robustum*) × *Miscanthus japonica*. Sugarcane Res Ann Prog Rep 3:377–386

Luo YW, Yen XC, Zhang GY, Liang GH (1992) Agronomic traits and chromosome behavior of autotetraploid sorghums. Plant Breed 109:46–53

Ma X-, Jensen E, Alexandrov N, Troukhan M, Zhang L, Thomas-Jones S, Farrar K, Clifton-Brown J, Donnison I, Swaller T, Flavell R (2012) High resolution genetic mapping by genome sequencing reveals genome duplication and tetraploid genetic structure of the diploid Miscanthus sinensis. PLoS ONE 7:e33821

Matumura M (1998a) Autecology of major forage grass (21): basic study for sustainable use. Anim Husbandry 52:717–725

Matumura M (1998b) Autecology of major forage grass (20): basic study for sustainable use. Anim Husbandry 52:627–634

Matumura M, Yukimura T (1975) Fundamental studies on artificial propagation by seeding useful wild grasses in Japan. VI. Germination behaviors of three native species of genus *Miscanthus*; *M. sacchariflorus*, *M. sinensis*, and *M. tinctorius*. Res Bull Fac Agric Gifu Univ 38:339–349

Matumura M, Hasegawa T, Saijoh Y (1985) Ecological aspects of *Miscanthus sinensis* var. *condensatus M. x sacchariflorus* and their 3x–4x-hybrids (1) Process of vegetative spread. Res Bull Fac Agric Gifu Univ 50:423–433

Matumura M, Hakumura Y, Saijoh Y (1986) Ecological aspects of *Miscanthus sinensis* var. *condensatus M.* × *sacchariflorus* and their 3x-4x-hybrids (2) Growth behaviour of the current year's rhizomes. Res Bull Fac Agric Gifu Univ 51:347–362

Matumura M, Hasegawa T, Saijoh Y (1987) Ecological aspects of *Miscanthus sinensis* var. *condensatus M. x sacchariflorus* and their 3x–4x-hybrids. (3) Aboveground standing crop and response to cutting. Res Bull Fac Agric Gifu Univ 52:315–324

Maximowicz M (1859) Primitae Florae Amurensis. Mem Acad Imp Sci St Pitersb 9:331

McNeill J, Barrie FR, Burdet HM, Demoulin V, Hawksworth DL, Marhold K, Nicolson DH, Prado J, Silva PC, Skog JE, Wiersema JH, Turland NJ (eds) (2006) International Code of Botanical Nomenclature (Vienna Code) adopted by the Seventeenth International Botanical Congress Vienna, Austria, July 2005, Regnum Vegetabile 146

Miyabuchi Y, Sugiyama S (2006) A 30,000-year phytolith record of a tephra sequence, east of Aso Caldera, southwestern Japan. Quater Res 45:15–28

Muntzing A (1951) Cytogenetic properties and practical value of tetraploid rye. Hereditas 37:17–84

Naidu SL, Moose SP, AL-Shoaibi AK, Raines CA, Long SP (2003) Cold tolerance of C_4 photosynthesis in *Miscanthus x giganteus*: Adaptation in amounts and sequence of C_4 photosynthetic enzymes. Plant Physiol 132:1688–1697

Newman M Ketphanh S Svengsuksa B Thoma P Sengdala K Lamxay V Armstrong K (2007) A Checklist of the vascular plants of Lao PDR. Royal Botanic Garden, Edinburgh, Scotland

Nielsen PN (1990) Elefantengrassanbau in Dänemark – Praktikerbericht. Pflug Spaten 3:1–4

Nimura M, Kato J, Horaguchi H, Mii M, Sakai K, Katoh T (2006) Induction of fertile amphidiploids by artificial chromosome-doubling in inter-specific hybrids between *Dianthus caryophyllus* L. and *D. japonicus* Thunb. Breed Sci 56:303–310

Nishiwaki A, Mizuguti A, Kuwabara S, Toma Y, Ishigaki G, Miyashita T, Yamada T, Matuura H, Yamaguchi S, Lane Rayburn A, Akashi R, Stewart RJ (2011) Discovery of natural *Miscanthus* (Poaceae) triploid plants in sympatric populations of *Miscanthus sacchariflorus* and *Miscanthus sinensis* in southern Japan. Am J Bot 98:154–159

Ogura J, Yamamoto S, Ikeka A (2002) The origin of the grassland of Aso region, Kyushu Japan, by microscopic charcoal analysis. Summaries Res AMS Nagoya Univ 13:236–240

Paterson AH, Schertz KF, Lin YA, Liu SC, Chang YL (1995) The weediness of wild plants: molecular analysis of genes influencing dispersal and persistence of johnsongrass, *Sorghum halepense* (L). Pers Proc Natl Acad Sci U S A 92:6127–6131

Pyter R, Voigt T, Heaton E, Dohleman F, Long S (2007) Giant Miscanthus: biomass crop for Illinois. In: Janick J, Whipkey A (eds) Issues in new crops and new uses 2007. ASHS, Alexandria, VA

Ramdoyal K, Badaloo GH (2002) Prebreeding in sugarcane with an emphasis on the programme of the Mauritius sugar industry research institute. In: Engles JMM, Rao VR, Brown AHD, Jackson MT (eds) Managing plant genetic diversity. IPGRI, Rome

Ramsey J, Schemske DW (1998) Pathways, mechanisms, and rates of polyploid formation in flowering plants. Ann Rev Ecol Syst 29:467–501

Rayburn AL, Crawford J, Rayburn CM, Juvik JA (2009) Genome size of three *Miscanthus* species. Plant Mol Biol Rep 27:184–188

Reeder J (1948) The Gramineae-Panicoideae of New Guinea. J Arnold Arbor 29:321–392

Renvoize SA (2003) The genus *Miscanthus*. Plantsman 2:207–211

Scally L, Wladren S, Hodkinson TR, Jones MB (2001) Morphological and molecular systematics of the genus *Miscanthus*. Aspects Appl Biol 65:231–237

Scurlock JMO (1998) *Miscanthus*: a review of European experience with a novel energy crop. ORNL/TM-13732. Oak Ridge National Laboratory, Oak Ridge, TN, 26 pp

Shouliang C, Renvoize SA (2006) *Miscanthus*. Flora Chin 22:581–583

Stewart JR, Toma YO, Fernandez FG, Nishiwaki A, Yamada T, Bollero GN (2009) The ecology and agronomy of Miscanthus sinensis a species important to bioenergy crop development in its native range in Japan: a review. GCB Bioenergy 1:126–153

Sun Q (2009) Primary taxonomic study of *Miscanthus* Andersson s.l. (Poaceae) from China and Japan. Dissertation Institute of Botany, the Chinese Academy of Sciences

Sun Q, Lin Q, Yi ZL, Yang ZR, Zhou F (2010) A taxonomic revision of *Miscanthus* Andersson s.l. (Poaceae) from China. Bot J Linn Soc 164:178–220

Swaminathan K, Alabady MS, Varala K, De Paoli E, Ho I, Rokhsar DS, Arumuganathan AK, Ming R, Green PJ, Meyers BC, Moose SP, Hudson ME (2010) Genomic and small RNA sequencing of *Miscanthus* × *giganteus* shows the utility of sorghum as a reference genome sequence for Andropogoneae grasses. Genome Biol 11:R12

Swaminathan K, Chae WB, Mitros T, Varala K, Xie L, Barling A, Glowacka K, Hall M, Jezowski S, Ming R, Hudson M, Juvik JA, Rokhsar DS, Moose SP (2012) A framework genetic map for *Miscanthus sinensis* from RNAseq-based markers shows recent tetraploidy. BMC Genomics 13:142–159

Taliaferro CM, Vogel KP, Bouton JH, McLaughlin SB, Tuskan GA (1999) Reproductive characteristics and breeding improvement potential of switchgrass. In: Biomass, a growth opportunity in green energy and value-added products, proceedings of the 4th biomass conference of the Americas, 29 August to 2 September

Thomas H (1993) Chromosome manipulation and polyploidy. In: Hayward M, Bosemark N, Romagosa I (eds) Plant breeding: principals and prospects. Chapman and Hall, London, pp 79–92

Tu S, Luan L, Liu Y, Long W, Kong F, He T, Xu Q, Yan W, Yu M (2007) Production and heterosis analysis of rice autotetraploid hybrids. Crop Sci 47:2356–2363

Ueda Y (1994) Systematic studies in the genus *Rosa*. Technol Bull Fac Hortic Chiba Univ Jpn 48:241–328

Wang D, Portis AR, Moose SP, Long SP (2008) Cool C_4 photosynthesis: pyruvate P_i dikinase expression and activity corresponds to the exceptional cold tolerance of carbon assimilation in *Miscanthus* × *giganteus*. Plant Physiol 148:557–567

Wang X, Yamada T, Kong F-J, Abe Y, Hoshino Y, Sato H, Takamizo T, Kanazawa A, Yamada T (2011) Establishment of an efficient in vitro culture and particle bombardment-mediated transformation systems in *Miscanthus sinensis* Anderss., a potential bioenergy crop. GCB Bioenergy 3:322–332

Watanabe H, Takahashi Y (2006) Dyeing golden by *Miscanthus tinctorius*. Bull Jpn Assoc Bot Gardens 40:81–87

Xi Q (2000) Investigation on the distribution and potential of giant grasses in China – *Triarrhena*, *Miscanthus*, *Arundo*, *Phragmites* and *Neyraudia*. Cuvillier, Goettingen

Xi Q (2003) Potential of Giant Grass Triarrhena lutarioriparia to grow in cold, dry and saline conditions as energy source. In: Proceedings of the International Conference on Bioenergy Utilization and Environment Protection - 6th LAMNET Project Workshop, 24–26 September, Dalian, China

Xiao FH, Tai PYP (1994) Antheral transformation into stigma in interspecific and intergeneric hybrids of *Saccharum*. J Am Soc Sugar Cane Technol 14:33–39

Yi ZL, Zhou PH, Chu CC, Li X, Tian WZ, Wang L, Cao SY, Tang ZS (2001) Establishment of genetic transformation system for *Miscanthus sacchariflorus* and obtained of its transgenic plant. Gaojishu Tongxin/High Technology Letters 11(4):20

Yoshida M, Liu Y, Uchida S et al (2008) Effects of cellulose crystallinity, hemicellulose, and lignin on the enzymatic hydrolysis of *Miscanthus sinensis* to monosaccharides. Biosci Biotechnol Biochem 72:805–810

Yu CY, Kim HS, Rayburn AL, Widholm JM, Juvik JA (2009) Chromosome doubling of the bioenergy crop *Miscanthus* × *giganteus*. GCB Bioenergy 1:404–412

Part II
Genomic Tools, Resources and Approaches

Chapter 5
The Sorghum Genome Sequence: A Core Resource for Saccharinae Genomics

Andrew H. Paterson

Abstract As a taxon noted for large and complex polyploid genomes, a facile genomic model is of especially great importance to the Saccharinae. The genome of *Sorghum bicolor* (sorghum) offers numerous advantages as such a model, with a physical size (about 730 mbp) that is only moderately larger than that of rice, and enjoying the same low level of gene duplication as rice by virtue of a lack of genome duplication for 70 million years. Saccharinae, especially sugarcane, researchers have long exploited comparative genomics to leverage the small and well-mapped sorghum genome in the study and improvement of more complex genomes. The sequencing of the sorghum genome further enhances such leveraging opportunities, also providing insights into genes and genomic features that may contribute to distinguishing features of the Saccharinae. A host of postgenomic tools for sorghum, many described elsewhere in this volume, provide the foundation for use of sorghum as a Saccharinae functional genomics model. Here, we revisit the sequencing and initial analysis of the sorghum genome, providing more detail than could be included in the primary description of the genome and also highlighting planned efforts to increase knowledge of sequence diversity in the species and the genus.

Keywords Whole-genome shotgun sequence • Repetitive DNA • Gene repertoire • Recombination • Genome duplication • Sequence diversity • Single-nucleotide polymorphism

A.H. Paterson (✉)
Plant Genome Mapping Laboratory, Departments of Crop and Soil Science, Plant Biology, and Genetics, University of Georgia, Athens, GA, USA
e-mail: paterson@plantbio.uga.edu; paterson@uga.edu

A.H. Paterson (ed.), *Genomics of the Saccharinae*, Plant Genetics and Genomics: Crops and Models 11, DOI 10.1007/978-1-4419-5947-8_5,

1 Rationale for Sorghum as a Saccharinae Botanical and Genomic Model

The small genome of sorghum has long been an attractive model for advancing understanding of the structure, function, and evolution of grass genomes. Sorghum is representative of many grasses of tropical origin in that it has "C4" photosynthesis, using complex biochemical and morphological specializations to improve carbon assimilation at high temperatures. By contrast, rice, the first grass genome to be sequenced, is more representative of temperate grasses, using "C3" photosynthesis. Sorghum and rice each share the distinction, as does Brachypodium (Initiative 2010) of a lack of genome duplication since an event in a common ancestor of the three about 70 mya (Paterson et al. 2004b). The resulting low levels of gene duplication make these three taxa attractive models for functional genomics. However, sorghum is much more closely related than rice to several major crops with complex genomes and high levels of gene duplication. *Sorghum* and *Zea* (maize, the leading US crop with a farm-gate value of $15–20 billion/year, in the Andropogonae but just outside the Saccharinae) diverged from a common ancestor ~12 mya (Gaut et al. 1997; Swigonova et al. 2004a) versus ~42 mya for rice and the maize/sorghum lineage (Paterson et al. 2004b). *Saccharum* (sugarcane), arguably the most important biofuels crop worldwide, valued at ~$30 billion including $1 billion/year in the US*, may have shared ancestry with sorghum as little as 5–9 mya (Sobral et al. 1994; Jannoo et al. 2007), retains similar gene order (Ming et al. 1998), and even produces viable progeny in some intergeneric crosses (Dewet et al. 1976). *Zea* has undergone one whole-genome duplication since its divergence from *Sorghum* (Swigonova et al. 2004b), and *Saccharum* has undergone at least two (Ming et al. 1998).

Many members of the Saccharinae group of cereals (see Chap. 1 by Kellogg, this volume) have large and complex genomes that derive great benefit from a closely related genomic model. This interesting group shows sixfold variation in genome size among closely related species with the same chromosome number (for example, *S. bicolor* and *S. propinquum* versus *S. nitidum*) (Price et al. 2005), an apparent reduction in chromosome number from the ancestral $2n = 20$ to $2n = 10$ in most parasorghums (Spangler et al. 1999), at least two chromosome doublings in *Saccharum* since its divergence from the remainder of the group (Ming et al. 1998), and both natural (*Sorghum halepense*: (Paterson et al. 1995)) and human-mediated polyploidization (*Saccharum* cultivars: (Ming et al. 1998)).

Miscanthus, a promising bioenergy crop for temperate latitudes (Heaton et al. 2008), is thought to have diverged from *Saccharum* more recently than their common ancestor did from sorghum. *Miscanthus* species have a basal set of 19 chromosomes ($2n = 38$, and 38 or 76 for spp. *sinensis* and *sacchariflorus*, respectively), versus the 10 that is characteristic of many Saccharinae including *Saccharum*. One attractive hypothesis to explain the transition from 10 to 19 chromosomes is that *Miscanthus*, like *Saccharum* (Ming et al. 1998), may have experienced a polyploidization in the 8–9 my since its divergence from sorghum. However, its basal chromosome number of 19 suggests that unlike those of *Saccharum* which is largely autopolyploid,

Miscanthus homologs may have diverged sufficiently that they no longer normally pair with one another (perhaps including a chromosomal fusion to get from 20 to 19). Perhaps *Miscanthus* and *Saccharum* even shared a genome doubling, diverging later. However, other options are plausible, such as "breakage" of *Miscanthus* chromosomes, or even some combination of polyploidy followed by chromosome number reduction as is evident in several *Sorghum* species and suggested in maize (Tang et al. 2008). Genetic mapping of *Miscanthus* by several groups promises to clarify the relationship of its chromosomes to those of sorghum and *Saccharum*. Knowledge of the mechanisms, levels and patterns of evolution of genome size and structure in this curious group will help to reveal the path by which the sorghum genome has arrived at its present state, also laying the foundation for further study of sugarcane and other economically important members of the group.

2 The Sorghum Genome Sequence

2.1 Sequencing and Assembly

Assembly of the sorghum genome (Paterson et al. 2009) was based on integration of a whole-genome shotgun sequence (Gardner et al. 1981) with a detailed genetic map (Chittenden et al. 1994; Bowers et al. 2003) and a BAC-based physical map (Bowers et al. 2005) that was rich in both BAC end sequences and sequence-tagged hybridization probes (many of which were derived from genetically mapped sequences).

The breeding line BTx623 was sequenced, a largely homozygous advanced breeding line released by Texas A&M University (Frederiksen and Miller 1972), which figures prominently in the pedigrees of many elite sorghum genotypes and has been widely used in sorghum genomics research. Approximately 8.5-fold redundant paired-end shotgun sequencing was performed using standard Sanger sequencing methodologies from small (~2–3 kb), and medium (5–8 kb) insert plasmid libraries, one fosmid library (~35 kb inserts), and two BAC libraries (average insert size 90 and 108 kb).

An initial whole genome shotgun (WGS) assembly was built with Arachne2 (Jaffe et al. 2003). The 201 largest WGS scaffolds, spanning 678.9 mbp and representing 97.3 % of the assembly, were scrutinized for discrepancies with genetic and/or physical maps as well as with the rice sequence. The genetic map was composed of 2,512 loci that defined 61.5 % of the recombination events in the underlying population (Bowers et al. 2003), with sequences available from 2,050 probes that identified about 90 % of the loci (some probes hybridizing to multiple loci). The physical map consisted of 1,869 contigs assembled from an 11× coverage BAC library by BAC fingerprinting, also including paired-end sequences for 96,870 BAC clones. The BAC contigs averaged 35.6 BACs in length, and in addition to BAC ends also contained an average of 10.1 hybridization loci, many derived from the genetically mapped probes (Bowers et al. 2005).

Iterative inspection of the initial sorghum sequence assembly, genetic and physical maps, and rice sequence at points of discrepancy led to the inference of a total of 28 putative assembly errors that were each supported by multiple lines of evidence and often involved repetitive elements. After breaking the WGS assembly at the 28 points of discrepancy with the genetic and physical maps, 127 of the resulting 229 scaffolds containing 625.7 mbp (89.7 %) of DNA could be assigned to chromosomal locations and oriented based on physical map, genetic map, rice synteny, genome structure (gene and repeat distributions), and cytological information (Kim et al. 2005). Most chromosomal models appeared largely complete: 15 of 20 terminated in telomeric repeats.

Two measures were taken to assess the completeness of the *S. bicolor* assembly. A total of 20,417 *S. bicolor* transcript assemblies from the TIGR PlantTA gene indices were aligned using BLAT (Kent 2002) to the repeat-masked sequence. Only 911, or 4.4 %, did not map to the genome assembly. Of these, only 51 were shown to have any similarity to known plant sequences, with the remainder dominated by hits to fungal genes (related to the genus *Fusarium*) or other likely contaminants of available sorghum cDNA libraries, suggesting that the assembly has missed fewer than 1 % of transcripts. Further, 31 BAC clones were subcloned into ~3 kb insert plasmid clones and end-sequenced using ABI3730 Sanger methods, and finished to Bermuda standards by primer walking and gap closure. Comparison of the assembly to these randomly chosen BAC clones showed that 98.46 % of the bases were represented in the assembly exactly as they appeared in the clones. The nonmatching bases were due largely to four assembly collapses on repetitive elements that account for 35,040 (~1 %) of the nonmatching bps in the 3.3 Mb surveyed, and one finished clone deletion of 4,223 bps.

The availability of both a sequence assembly and a BAC-based physical map permitted comparison of the relative quality of these two independent data types. Physical maps are prone to a variety of assembly artifacts, relating to such factors as uneven coverage of the genome by BAC libraries, specific band sizes that are abundant in the genome as a result of sequence repetition, and low complexity of BACs in repetitive domains of a genome. A total of 38 (2 %) of 1,869 FPC contigs from the *S. bicolor* physical map(Bowers et al. 2005) were considered erroneous, based on containing >5 BAC-ends that fell into different sequence scaffolds. In other words, the BAC-based physical map and the initial sequence assembly each had similar error rates, but the availability of each of these complementary resources contributed to the high quality of the sequence assembly.

2.2 Sorghum Gene Repertoire

Protein-coding genes were inferred from the consensus of several sources of evidence as well as ab initio predictions. First, TIGR rice transcript assemblies (Childs et al. 2007) were mapped to the repeat-masked Sorghum genome sequences applying

GenomeThreader (Gremme et al. 2005) and a maize splice site model. Optimal spliced alignments (OSAs) of assemblies and ESTs of the following monocot species were included: *Allium cepa*, *Ananas comosus*, *Avena sativa*, *Brachypodium distachyon*, *Curcuma longa*, *Hordeum vulgare*, *Oryza sativa*, *Saccharum officinarum*, *Secale cereale*, *Sorghum bicolor*, *Sorghum halepense*, *Sorghum propinquum*, *Triticum aestivum*, *Zea mays*, and *Zingiber officinale*. OSAs as well as BlastX alignments were also generated for a reference set of proteins consisting of the SWISSPROT database (Bairoch 2005) and proteomes of *Arabidopsis thaliana* (TAIR6 version; (Swarbreck et al. 2008)), *Saccharomyces cerevisiae* (Mewes et al. 2004), and rice (Tanaka et al. 2008). For each OSA, possible reading frames of size ≥50 amino acids were collected as candidates for gene models. In addition, gene models were identified on repeat masked genomic sequences by ab initio methods (Fgenesh++, GeneID, GenomeScan). Jigsaw (Allen and Salzberg 2005) was applied as a statistical combiner of all supporting information from this first analysis, using a decision tree trained on a set of 987 gene models that were edited by human supervision in the Apollo Genome Browser (Searle et al. 2004). All models, including those obtained from the first analysis series, were scored by Blastp against the UniREF90 protein database and for each locus the best fitting model, i.e., the model with the highest bit score, was used as input for the PASA pipeline (Haas et al. 2008) to (1) predict UTRs using maize, sorghum, and sugarcane ESTs, (2) identify possible alternative splicing patterns, and (3) fit all predicted models to the splice sites suggested by EST evidences of closely related species. Besides complete gene models, we also included candidate (partial) genes that lack a start and/or stop codon and may result from (1) sequencing or assembly errors, (2) transposon activity that may have led to truncated genes or pseudogenes, (3) insufficient evidence from ab initio predictions or EST matches leading to incomplete gene models.

We adopted a gene nomenclature convention used by the Arabidopsis and rice communities, with each protein-coding gene locus assigned a unique identifier of the form "SbXX%YYYYY" where:

- "Sb" indicates *Sorghum bicolor*.
- "XX" is a two digit numerical chromosome identifier (01-10) or four digit scaffold identifier (0010-3326).
- The delimiter "%" is either "g" for chromosomally mapped sequences or "s" for scaffolds.
- YYYYY is a unique five digit numerical code, starting from 00200 at the start of each assembled sequence, and incrementing by ten. Spans longer than 100 kb between initially annotated loci are represented by a skip of 200. Thus, in the initial assignment, the numerical code corresponds to chromosomal position, but leaving flexibility to add new genes as discovered while preserving the identifiers of already-discovered genes.

Among 34,496 sorghum gene models inferred by this pipeline (Paterson et al. 2009), ~27,640 were supported by both homology-based and ab initio gene prediction methods with extensive public expressed sequences from sorghum, maize, and

sugarcane, and were deemed "bona fide protein-coding genes." Evidence for alternate splicing is found in 1,491 loci. In addition to these high confidence protein-coding genes, another 5,197 predicted gene models were generally shorter than the bona fide genes (often <150 amino acids); had few exons (often one) and no EST support; were more extensively diverged from related rice genes; and were often found in large families enriched for "hypothetical," "uncharacterized," and/or retroelement-associated domains and annotations. Relatively high concentration of these low confidence gene models in the pericentromeric regions where bona fide genes are scarce suggests that many are retroelement-derived. A total of 727 processed pseudogenes (intronless copies of intron-containing genes elsewhere in the genome) and 932 predictions containing domains only known from transposable elements were identified.

A total of 29,734 protein functional domains (nonredundant, i.e., counting a PFAM domain only once per gene) were detected in 21,743 sorghum genes. The number and sizes of sorghum gene families were similar to those for *Arabidopsis*, rice, and poplar. Several domains that are overrepresented, underrepresented, or even absent in sorghum relative to rice, poplar, and *Arabidopsis*, may represent molecular fingerprints of biological peculiarities specific to the *Sorghum* lineage (see Paterson Chap. 23, this volume).

A total of 19,929 sorghum genes had syntenic orthologs in the available rice annotations, the manually curated and expressed-sequence focused "RAP2" annotation from the Rice Annotation Project (Tanaka et al. 2008) and the automated "TIGR5" annotation from TIGR (including co-orthologs arising from local tandem duplications in the sorghum and/or rice genomes). An additional 3,919 sorghum genes have a mutual best (or near-best) BLAST match to a rice gene that is not in a corresponding syntenic location but are similar to syntenic orthologs in their lengths, synonymous substitution rates, and introns, and are presumed to be orthologous genes whose chromosomal positions have changed in one or both grass lineages.

Most paralogs in the sorghum genome are members of tandem arrays, with 5,303 genes in 1,947 families of two or more proximally duplicated genes. Tandem expansions were defined as all sets of peptides with a pairwise Blastp alignment of *e*-value better than 1*e*−25 and two or less intervening genes. The longest tandem gene array is composed of 15 cytochrome P450 genes (first gene Sb03g028560.1), with some other long arrays including GRAS family transcription factors (14: Sb05g027740.1), Chalcone and stilbene synthases (14: Sb05g019890.1), an auxin-responsive protein (14: Sb02g031700.1), homologs of an Arabidopsis protein of unknown function (13: Sb07g024600.1), a BTB/POZ containing family (13: Sb07g026660.1), a leucine-rich repeat containing family (12: Sb04g003800.1), and glutathione S-transferase (12: Sb01g030930.1). Some gene families tandemly amplified in sorghum contain motifs associated with transposable elements (Zinc knuckle, haT family dimerization domain, reverse transcriptase, transposase family *Tnp2*, transposase DDE domain, and putative gypsy type transposons), and may have escaped repeat masking or evolved from transposable elements by neofunctionalization, the latter consistent with their presence in the manually annotated rice genome.

2.3 *Nature, Quantity, and Organization of Sorghum Repetitive DNA*

Recent renaturation kinetics studies show that the sorghum genome is largely comprised of repetitive DNA (Peterson et al. 2002b). Known repeats were identified with RepeatMasker (www.repeatmasker.org) with a database of previously known repeats from sorghum and other grasses (mips-REdat_6.2_Poaceae.lib).

LTR retrotransposons comprised 55 % of the sorghum genome, and about 98 % of the total repetitive DNA in the genome. De novo searches for performed with LTR_STRUCT (pmid 12584121) on the ten sorghum chromosomes and all unassembled contigs >10 kb yielded 10,126 full-length LTR retrotransposon candidate sequences. A total of 8,071 (80 %) of the candidate sequences were nonoverlapping, contained at least one typical retrotransposon protein domain GAG, PR, INT, RT and had a simple sequence and tandem repeat content <=35 %. According to their protein signatures 2,985 (37 %) could be assigned to the gypsy (PR-RT-INT) and 724 (9 %) to the copia (PR-INT-RT) LTR superfamily, the remaining 4,362 (54 %) remaining initially unclassified. A nonredundant set of 7,643 quality checked LTR retrotransposons was added to mipsREdat (mips.gsf.de/proj/plant/webapp/recat/), a plant repeat element database, used for the homology based repeat masking and annotation. The single most abundant element family alone, Retrosor-6 (Peterson et al. 2002a), comprises ~6.9 % of the genome.

LTR retroelements of different ages showed striking differences in genomic distribution (Paterson et al. 2009). The insertion age of full length LTR-retrotransposons was determined from the evolutionary distance between 5′ and 3′ soloLTR derived from a ClustalW alignment of the two solo LTRs by the Kimura two-parameter method (emboss distmat, http://emboss.sourceforge.net/), using a substitution rate of $1.3e{-}8$ mutations per site per year. Relatively young (i.e., less than 0.1 mya) LTR insertions are approximately uniformly distributed across the genome, while older insertions are heavily concentrated in heterochromatic regions immediately surrounding the centromeres. This closely parallels a proposal based on comparison of the sorghum physical map to the rice sequence, that selection against gene and genome structural mutation is much stronger in the gene-rich euchromatin than in the repeat-rich heterochromatin (Bowers et al. 2005).

DNA transposons constitute about 7.5 % of the sorghum genome. CACTA-like elements, the predominant class of DNA transposon in sorghum (4.7 % of the genome), also appear to be the predominant mechanism for translocation of sorghum genes and gene fragments. Most CACTA elements have arrays of direct and inverted repeat units (about 20-40 bp per unit) in their terminal regions, and could be identified by this feature and the characteristic palindromic CACTA/G … C/TAGTG termini flanked by a 3 bp target site duplication. Among 95 novel CACTA families discovered in the sorghum sequence, most individual elements were nonautonomous deletion derivatives of 5–7 kb in which the typical transposon genes have been replaced with nontransposon DNA including exons from one or more genes.

For example, CACTA family *G118* has only one complete and presumably autonomous "mother" element, with 8 of 18 deletion derivatives carrying gene fragments internally. Among the 13,775 CACTA elements identified, 200 encode no transposon proteins but contain at least one fragment of a cellular gene. Thus, the sorghum lineage shares the gene-transducing property that has been observed in other cereal genomes, although "utilizing" a different transposon family than the Mutator-like "Pack-MULE" elements that are the predominant gene-transducing element(Jiang et al. 2004) in rice, and intact helitrons that are implicated in maize gene movement(Brunner et al. 2005).

Miniature inverted-repeat transposable elements (MITEs) account for most of the non-CACTA DNA transposons, comprising 1.7 % of the sorghum genome. Full-length MITEs were identified based on their inverted repeat structure and their 2 and 3 bp target site duplications for *Stowaway* and *Tourist* MITEs, respectively. The initial set identified in this way was used for multiple sequence alignments in order to identify families and to construct consensus sequences for all of them. The consensus sequences were used for a BLAST survey to identify all elements, including fragments.

Tandemly repeated DNA was closely associated with reduced power to assemble both genomic sequence and physical BAC contigs. After breaking the WGS assembly at the 28 points of discrepancy with the genetic and physical maps (see above), 127 of the resulting 229 scaffolds containing 625.7 mbp (89.7 %) of DNA could be assigned to chromosomal locations and oriented based on physical map, genetic map, rice synteny, genome structure (gene and repeat distributions), and cytological information (Kim et al. 2005). These 127 scaffolds merged 1,476 FPC contigs. The remaining 102 scaffolds were generally smaller (53.2 mbp, 7.6 % of nucleotides), with only 374 predicted genes and 85 (83 %) containing large stretches comprised predominantly of the CEN38 (Miller et al. 1998) centromeric repeat. These 102 scaffolds merged only 193 physical BAC contigs, due to the greater abundance of repeats that are recalcitrant to clone-based physical mapping (Bowers et al. 2005) and may be omitted in BAC-by-BAC approaches (Venter et al. 1998). In a like manner, it was previously shown (Bowers et al. 2005) that singleton sorghum BACs that could not be joined to FPC assemblies were gene-poor (3.9 % of one-locus and 4 % of ten-locus probe anchors), but rich in centromeric repeats (19–23 % of all centromeric anchored BACs), with substantially fewer fingerprint bands than contigged BACs, also consistent with being largely composed of tandem repeats with multiple copies of the same band size.

2.4 Organellar DNA

The use of an efficient nuclear DNA isolation method in preparation of DNA for sequencing left insufficient organelle "contamination" in the shotgun data to recreate both organelles from the WGS set. However, comparison of the assembled nuclear genome of Sorghum to the Sorghum plastid genome (EF115542) and the

sorghum mitochondrial genome (DQ984518) respectively, revealed numerous insertions of DNA from mitochondria (2,125 insertions comprising 305,381 bp, or 0.046 % of the genome) and chloroplast (1,402 insertions, 255,600 bp: 0.039 %), in the nuclear genome sequence. While numerous, these insertions accounted for a much smaller portion of the genome than organellar insertions in rice (0.29 and 0.24 %, respectively), with the difference much larger than could be attributed to LTR retroelement expansion in sorghum. Most insertions (96 and 95 %) are <500 bp, with only 1.5 % being >2 kb. Longer insertions show more sequence conservation with organellar DNA, suggesting that they are generally more recent—this implies that longer insertions are more prone to removal than short insertions.

3 Sorghum Genome Organization, Transmission, and Evolution

Remarkably, about 62 % of the sorghum genome is almost recombinationally inert. A genetic map based on a cross of BTx623 (the sequenced genotype) and its wild relative *S. propinquum* (also physically mapped; (Bowers et al. 2005)) including 2,512 loci that defined 61.5 % of the recombination events in the underlying population (Bowers et al. 2003), could be aligned with the genome sequence based on sequences for 2,050 probes that identified about 90 % of the loci (some probes hybridizing to multiple loci). The genome could be clearly partitioned into a recombination-rich fraction estimated to comprise 252 mbp of DNA and accounting for 1,025.2 cM (97 %) of the length of the genetic map; and a recombination-poor fraction comprising about 460 mbp of DNA and accounting for 34 cM (3 %) of the length of the genetic map.

Cytologically identified heterochromatin (Chen et al. 2002; Jiao et al. 2005) corresponds very closely to the recombinationally inert region(s) of the sorghum genome. The locations of heterochromatin in the sorghum physical map (Bowers et al. 2005) and sequence (Paterson et al. 2009) were estimated based on relative distance from the centromere, assigning approximate base pair locations without accounting for sequence gaps. The euchromatin–heterochromatin distinction is associated closely with not only differences in recombination rate, but also with differential abundance of low copy-number genes and gene-associated DNA transposons (frequent in euchromatin), and high-copy number genes and relatively "old" LTR retroelements (frequent in heterochromatin).

3.1 Intragenomic Duplication in Sorghum

Analysis of the sorghum genome sequence corroborated early suggestions based on genetic mapping that sorghum may be a paleopolyploid. RFLP mapping revealed

several instances in which groups of probes that segregated for multiple informative polymorphisms showed parallel organization in two different regions of the genome (Chittenden et al. 1994). Comprehensive comparison of the sorghum genes to one another (Paterson et al. 2009), revealed that such patterns of parallel organization could collectively account for the vast majority of the genome, in much the same manner as had been previously shown for the first sequenced grass genome, rice (Goff et al. 2002). Indeed, patterns of sequence correspondence in rice and sorghum were very similar. This was no surprise—sequence diversity between duplicated rice genes was known to be generally greater than between rice-sorghum orthologs, suggesting that the most recent genome duplication in each of these grasses had occurred in a common ancestor prior to the divergence of their respective lineages (Paterson et al. 2004b).

The genomic distribution of regions in which paleoduplication could/could not be discerned added further weight to the notion that the sorghum genome is comprised of two highly divergent "components" with remarkably different properties and evolutionary histories. Discernible paleoduplication was almost completely confined to the recombination-rich and gene-rich "euchromatin," being undetectable in the heterochromatin. This is consistent with the hypothesis (Bowers et al. 2005) that genomic rearrangements are usually deleterious, thus more likely to persist in nonrecombinogenic regions by virtue of "Muller's ratchet" (Muller 1964). Moreover, selection may actually favor the rapid restructuring of heterochromatic regions that contain centromeres (Bowers et al. 2005). A high concentration of rice genes duplicated by ancient polyploidy fall near K_s 0.85 (Paterson et al. 2004a). Rice gene pairs with K_s 0.2–0.6 tend to be concentrated in peri/centromeric regions, suggesting that a substantial restructuring of centromeric regions began shortly after polyploidization and lasted until about 16 mya. The corresponding region of sorghum also shows massive recombination suppression—since restructuring of the rice heterochromatin appears to have continued long after the rice–sorghum divergence (K_s 0.51: 1), one would predict that parallel restructuring may have occurred in sorghum.

Restructuring of centromeric regions after ancient polyploidization may perhaps have been under selection (Bowers et al. 2005), rather than merely being a passive consequence of Muller's ratchet (Muller 1932). In a newly formed autopolyploid [arguably the most frequent type (Harlan and Dewet 1975)] with sets of four homologous chromosomes, divergence of centromeric DNA may accelerate the transition to diploid inheritance, allowing more rapid allele frequency changes and reduced genetic load.

4 Sorghum Sequence Diversity

With a high-quality draft sequence of the first sorghum genome published, members of the sorghum community have turned their attention to a number of postgenomic activities, many of which are described in other chapters of this volume. Among

these activities is to learn about the levels and patterns of DNA-level variation in the *S. bicolor* species and its relatives, providing an essential framework toward the long-term goal of identifying and interpreting the complete repertoire of functional DNA-level variation in sorghum. Population-based methods for mapping complex traits, originally developed in the human genetics community, are now being applied to crop plants, and have been successful in analysis of previously identified candidate genes (Thornsberry et al. 2001; Whitt et al. 2002; Yu et al. 2005). Sorghum is well suited to genome-wide association mapping methods for identification of quantitative trait loci (QTL), with hundreds of QTL mapped (see many other chapters of this volume), several diversity panels available, and nested association mapping populations under development as detailed in Chap. 9 by Yu et al. (this volume).

With the completion of the sorghum genome sequence, *the major limitation in undertaking this work is the lack of sufficient markers to provide dense genome-wide coverage*. Rapid advances in SNP technology will soon permit a transition from indirect study of diagnostic markers to comprehensive SNP-based assessment of the entire transcriptome for functional variation. This will have numerous impacts, for example in applied sciences providing for a transition from indirect study of diagnostic markers to comprehensive SNP-based assessment of the entire transcriptome for functional variation. This framework of functional diversity information will also permit us to investigate fundamental questions of seminal importance to plant domestication and biomass productivity

Several resequencing studies of specific genotypes are now known to be underway at a number of institutions, toward specific applied goals. In addition, an international group has been successful in securing a portion of the sequencing capacity of the US Department of Energy Joint Genome Institute for resequencing of a carefully selected group of 38 genotypes that broadly sample the spectrum of genetic variation in sorghum and its immediate relatives. Specifically, whole-genome Illumina resequencing is planned for:

(1) *S. versicolor* and *S. propinquum*, divergent wild relatives that represent phylogenetic outgroups for cultivated sorghum. The latter was also the parent of the interspecific reference genetic map (Bowers et al. 2003), and has been physically mapped (Chittenden et al. 1994; Bowers et al. 2005).
(2) *S. bicolor* genotype IS3620C, that is highly divergent from the genotype used for sequencing (BTx623), which was the alternative parent of the intraspecific reference map (Xu et al. 1994) and has also been used in BAC library construction and physical mapping (Klein et al. 2000).
(3) *S. bicolor* genotype "Rio," representing "sweet sorghums" that are one promising avenue for biofuel production from sorghum.

Further, resequencing of cDNA pooled from multiple diverse tissues will be conducted for 34 additional genotypes that broadly sample the recognized botanical races of sorghum as well as wild germplasm. These genotypes collectively address a wide range of applied needs, for example, including a source of aluminum tolerance, key breeding lines from several countries, and a number of parents of crosses planned in a pan-African "Challenge Initiative" being coordinated by the CGIAR

Generation Challenge Program. In addition, broad sampling of wild sorghums will permit us to investigate changes associated with domestication and improvement at levels ranging from specific nucleotides to gene families and pathways. As of this writing, these genotypes are at various stages in the sequencing process, with the collective data expected to be completed sometime in 2011.

Acknowledgments The author thanks numerous colleagues and collaborators for valuable discussions and innumerable contributions to aspects of 17 years of background work that led to sequencing of the sorghum genome, the Joint Genome Institute of the US Department of Energy for doing the vast majority of sequencing and contributing greatly to annotation and analysis, and a long list of coauthors of the primary sequence description (cited above) whose work is reviewed here.

References

Allen JE, Salzberg SL (2005) JIGSAW: integration of multiple sources of evidence for gene prediction. Bioinformatics 21:3596–3603

Bairoch A (2005) From sequences to knowledge, the role of the Swiss-Prot component of UniProt. Mol Cell Proteomics 4:S2–S2

Bowers JE, Abbey C, Anderson S, Chang C, Draye X, Hoppe AH, Jessup R, Lemke C, Lennington J, Li Z, Lin YR, Liu SC, Luo L, Marler BS, Ming R, Mitchell SE, Kresovich S, Schertz KF, Paterson AH (2003) A high-density genetic recombination map of sequence-tagged sites for sorghum, as a framework for comparative structural and evolutionary genomics of tropical grains and grasses. Genetics 165:367–386

Bowers JE, Arias MA, Asher R, Avise JA, Ball RT, Brewer GA, Buss RW, Chen AH, Edwards TM, Estill JC, Exum HE, Goff VH, Herrick KL, Steele CLJ, Karunakaran S, Lafayette GK, Lemke C, Marler BS, Masters SL, McMillan JM, Nelson LK, Newsome GA, Nwakanma CC, Odeh RN, Phelps CA, Rarick EA, Rogers CJ, Ryan SP, Slaughter KA, Soderlund CA, Tang HB, Wing RA, Paterson AH (2005) Comparative physical mapping links conservation of microsynteny to chromosome structure and recombination in grasses. Proc Natl Acad Sci U S A 102:13206–13211

Brunner S, Fengler K, Morgante M, Tingey S, Rafalski A (2005) Evolution of DNA sequence nonhomologies among maize inbreds. Plant Cell 17:343–360

Chen MS, Presting G, Barbazuk WB, Goicoechea JL, Blackmon B, Fang FC, Kim H, Frisch D, Yu YS, Sun SH, Higingbottom S, Phimphilai J, Phimphilai D, Thurmond S, Gaudette B, Li P, Liu JD, Hatfield J, Main D, Farrar K, Henderson C, Barnett L, Costa R, Williams B, Walser S, Atkins M, Hall C, Budiman MA, Tomkins JP, Luo MZ, Bancroft I, Salse J, Regad F, Mohapatra T, Singh NK, Tyagi AK, Soderlund C, Dean RA, Wing RA (2002) An integrated physical and genetic map of the rice genome. Plant Cell 14:537–545

Childs KL, Hamilton JP, Zhu W, Ly E, Cheung F, Wu H, Rabinowicz PD, Town CD, Buell CR, Chan AP (2007) The TIGR plant transcript assemblies database. Nucleic Acids Res 35:D846–D851

Chittenden LM, Schertz KF, Lin YR, Wing RA, Paterson AH (1994) A detailed Rflp map of Sorghum-bicolor × S-propinquum, suitable for high-density mapping, suggests ancestral duplication of sorghum chromosomes or chromosomal segments. Theor Appl Genet 87:925–933

Dewet JMJ, Gupta SC, Harlan JR, Grassl CO (1976) Cytogenetics of introgression from Saccharum into Sorghum. Crop Sci 16:568–572

Frederiksen RA, Miller FR (1972) Proposal for release and increase ATx622, BTx622, ATx623, BTx623, ATx624, BTx624. Seed Release Committee of the Texas Agricultural Experiment Station Texas A&M, College Station, TX

Gardner RC, Howarth AJ, Hahn P, Brownluedi M, Shepherd RJ, Messing J (1981) The complete nucleotide-sequence of an infectious clone of cauliflower mosaic-virus by M13MP7 shotgun sequencing. Nucleic Acids Res 9:2871–2888

Gaut BS, Clark LG, Wendel JF, Muse SV (1997) Comparisons of the molecular evolutionary process at rbcL and ndhF in the grass family (Poaceae). Mol Biol Evol 14:769–777

Goff SA, Ricke D, Lan TH, Presting G, Wang RL, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchinson D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong JP, Miguel T, Paszkowski U, Zhang SP, Colbert M, Sun WL, Chen LL, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu YS, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T, Oliphant A, Briggs S (2002) A draft sequence of the rice genome (Oryza sativa L. ssp japonica). Science 296:92–100

Gremme G, Brendel V, Sparks ME, Kurtz S (2005) Engineering a software tool for gene structure prediction in higher organisms. Inform Software Technol 47:965–978

Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR (2008) Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. Genome Biol 9:R7

Harlan JR, Dewet JMJ (1975) Winge, O and Prayer, a—origins of polyploidy. Bot Rev 41:361–390

Heaton EA, Dohleman FG, Long SP (2008) Meeting US biofuel goals with less land: the potential of Miscanthus. Glob Chang Biol 14:2000–2014

Initiative TIB (2010) Genome sequencing and analysis of the model grass Brachypodium distachyon. Nature 463:763–768

Jaffe DB, Butler J, Gnerre S, Mauceli E, Lindblad-Toh K, Mesirov JP, Zody MC, Lander ES (2003) Whole-genome sequence assembly for mammalian genomes: Arachne 2. Genome Res 13:91–96

Jannoo N, Grivet L, Chantret N, Garsmeur O, Glaszmann JC, Arruda P, D'Hont A (2007) Orthologous comparison in a gene-rich region among grasses reveals stability in the sugarcane polyploid genome. Plant J 50:574–585

Jiang N, Bao ZR, Zhang XY, Eddy SR, Wessler SR (2004) Pack-MULE transposable elements mediate gene evolution in plants. Nature 431:569–573

Jiao Y, Jia P, Wang X, Su N, Yu S, Zhang D, Ma L, Feng Q, Jin Z, Li L, Xue Y, Cheng Z, Zhao H, Han B, Deng XW (2005) A tiling microarray expression analysis of rice chromosome 4 suggests a chromosome-level regulation of transcription. Plant Cell 17:1641–1657

Kent WJ (2002) BLAT—the BLAST-like alignment tool. Genome Res 4:656–664

Kim JS, Klein PE, Klein RR, Price HJ, Mullet JE, Stelly DM (2005) Chromosome identification and nomenclature of Sorghum bicolor. Genetics 169:1169–1173

Klein PE, Klein RR, Cartinhour SW, Ulanch PE, Dong JM, Obert JA, Morishige DT, Schlueter SD, Childs KL, Ale M, Mullet JE (2000) A high-throughput AFLP-based method for constructing integrated genetic and physical maps: progress toward a sorghum genome map. Genome Res 10:789–807

Mewes HW, Amid C, Arnold R, Frishman D, Guldener U, Mannhaupt G, Munsterkotter M, Pagel P, Strack N, Stumpflen V, Warfsmann J, Ruepp A (2004) MIPS: analysis and annotation of proteins from whole genomes. Nucleic Acids Res 32:D41–D44

Miller JT, Jackson SA, Nasuda S, Gill BS, Wing RA, Jiang J (1998) Cloning and characterization of a centromere-specific repetitive DNA element from Sorghum bicolor. Theor Appl Genet 96:832–839

Ming R, Liu SC, Lin YR, da Silva J, Wilson W, Braga D, van Deynze A, Wenslaff TF, Wu KK, Moore PH, Burnquist W, Sorrells ME, Irvine JE, Paterson AH (1998) Detailed alignment of Saccharum and Sorghum chromosomes: comparative organization of closely related diploid and polyploid genomes. Genetics 150:1663–1682

Muller HJ (1932) Some genetic aspects of sex. Am Nat 66:118–138

Muller HJ (1964) The relation of recombination to mutational advance. Mutat Res 1:2–9

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti

AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob ur R, Ware D, Westhoff P, Mayer KF, Messing J, Rokhsar DS (2009) The Sorghum bicolor genome and the diversification of grasses. Nature 457:551–556

Paterson AH, Bowers JE, Chapman BA (2004a) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc Natl Acad Sci U S A 101:9903–9908

Paterson AH, Bowers JE, Chapman BA (2004b) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc Natl Acad Sci U S A 101:9903–9908

Paterson AH, Schertz KF, Lin YR, Liu SC, Chang YL (1995) The weediness of wild plants—molecular analysis of genes influencing dispersal and persistence of Johnsongrass, Sorghum halepense (L) Pers. Proc Natl Acad Sci U S A 92:6127–6131

Peterson DG, Schulze SR, Sciara EB, Lee SA, Bowers JE, Nagel A, Jiang N, Tibbitts DC, Wessler SR, Paterson AH (2002a) Integration of Cot analysis, DNA cloning, and high-throughput sequencing facilitates genome characterization and gene discovery. Genome Res 12:795–807

Peterson DG, Schulze SR, Sciara EB, Lee SA, Bowers JE, Nagel A, Jiang N, Tibbitts DC, Wessler SR, Paterson AH (2002b) Integration of Cot analysis, DNA cloning, and high-throughput sequencing facilitates genome characterization and gene discovery. Genome Res 12:795–807

Price HJ, Dillon SL, Hodnett G, Rooney WL, Ross L, Johnston JS (2005) Genome evolution in the genus Sorghum (Poaceae). Ann Bot 95:219–227

Searle SMJ, Gilbert J, Iyer V, Clamp M (2004) The Otter annotation system. Genome Res 14:963–970

Sobral BWS, Braga DPV, Lahood ES, Keim P (1994) Phylogenetic analysis of chloroplast restriction enzyme site mutations in the Saccharinae Griseb subtribe of the Andropogoneae Dumort tribe. Theor Appl Genet 87:843–853

Spangler R, Zaitchik B, Russo E, Kellogg E (1999) Andropogoneae evolution and generic limits in Sorghum (Poaceae) using ndhF sequences. Syst Bot 24:267–281

Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hernandez M, Foerster H, Li D, Meyer T, Muller R, Ploetz L, Radenbaugh A, Singh S, Swing V, Tissier C, Zhang P, Huala E (2008) The Arabidopsis Information Resource (TAIR): gene structure and function annotation. Nucleic Acids Res 36:D1009–D1014

Swigonova Z, Lai J, Ma J, Ramakrishna W, Llaca V, Bennetzen JL, Messing J (2004a) Close split of sorghum and maize genome progenitors. Genome Res 14:1916–1923

Swigonova Z, Lai JS, Ma JX, Ramakrishna W, Llaca M, Bennetzen JL, Messing J (2004b) On the tetraploid origin of the maize genome. Comp Funct Genomics 5:281–284

Tanaka T, Antonio BA, Kikuchi S, Matsumoto T, Nagamura Y, Numa H, Sakai H, Wu J, Itoh T, Sasaki T, Aono R, Fujii Y, Habara T, Harada E, Kanno M, Kawahara Y, Kawashima H, Kubooka H, Matsuya A, Nakaoka H, Saichi N, Sanbonmatsu R, Sato Y, Shinso Y, Suzuki M, Takeda JI, Tanino M, Todokoro F, Yamaguchi K, Yamamoto N, Yamasaki C, Imanishi T, Okido T, Tada M, Ikeo K, Tateno Y, Gojobori T, Lin YC, Wei FJ, Hsing YI, Zhao Q, Han B, Kramer MR, McCombie RW, Lonsdale D, O'Donovan CC, Whitfield EJ, Apweiler R, Koyanagi KO, Khurana JP, Raghuvanshi S, Singh NK, Tyagi AK, Haberer G, Fujisawa M, Hosokawa S, Ito Y, Ikawa H, Shibata M, Yamamoto M, Bruskiewich RM, Hoen DR, Bureau TE, Namiki N, Ohyanagi H, Sakai Y, Nobushima S, Sakata K, Barrero RA, Sato Y, Souvorov A, Smith-White B, Tatusova T, An S, An G, Oota S, Fuks G, Messing J, Christie KR, Lieberherr D, Kim H, Zuccolo A, Wing RA, Nobuta K, Green PJ, Lu C, Meyers BC, Chaparro C, Piegu B, Panaud O, Echeverria M (2008) The rice annotation project database (RAP-DB): 2008 update. Nucleic Acids Res 36:D1028–D1033

Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH (2008) Synteny and colinearity in plant genomes. Science 320:486–488

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES (2001) Dwarf8 polymorphisms associate with variation in flowering time. Nat Genet 28:286–289

Venter JC, Adams MD, Sutton GG, Kerlavage AR, Smith HO, Hunkapiller M (1998) Shotgun sequencing of the human genome. Science 280:1540–1542

Whitt SR, Wilson LM, Tenaillon MI, Gaut BS, Buckler ES (2002) Genetic diversity and selection in the maize starch pathway. Proc Natl Acad Sci U S A 99:12959–12962

Xu GW, Magill CW, Schertz KF, Hart GE (1994) A Rflp linkage map of Sorghum-Bicolor (L) Moench. Theor Appl Genet 89:139–145

Yu J, Pressoir G, Briggs W, Bi IV, Yamasaki M, Doebley J, McMullen M, Gaut B, Nielsen D, Holland J, Kresovich S, Buckler E (2005) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208

Chapter 6
Transcriptome Analysis in the Saccharinae

Milton Yutaka Nishiyama-Jr, Fabio Vicente, Paloma Mieko Sato, Savio Siqueira Ferreira, Frank Alex Feltus, and Glaucia Mendes Souza

Abstract Transcriptomics data for the Saccharinae clade are presently derived from EST projects for Saccharum, Sorghum and Miscanthus, with little known about other members of this clade. Among sugarcane EST data sets, by far the largest and most detailed is from the Brazilian SUCEST project. Rich EST resources are also available for Sorghum, although Miscanthus EST resources remain limited. The complete sequencing of the Sorghum genome led to a great improvement in this field, but the complexity of the sugarcane genome and lack of a genome sequence remain a hindrance. Comparative analysis of sorghum and sugarcane reveals a high general colinearity and gene structure conservation. Several platforms provide data on different traits and/or treatments such as brix accumulation, biotic and abiotic stress (cold and drought, for example) and tissue specificity of gene expression, shedding early light on how the control/response to these processes occurs at the gene expression level.

Keywords Sugarcane • Sorghum • Miscanthus • Microarray • Expression profile • EST database • Saccharinae comparison

M.Y. Nishiyama-Jr. (✉) • F. Vicente • P.M. Sato • S.S. Ferreira • G.M. Souza
Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo,
Avenida Lineu Prestes, 748 Bloco 9 Superior Sala 954, São Paulo - SP, CEP 05508-000, Brazil
e-mail: yutakajr@iq.usp.br; glmsouza@iq.usp.br

F.A. Feltus
Department of Genetics and Biochemistry, Clemson University,
Clemson, SC, USA
e-mail: FFELTUS@clemson.edu

A.H. Paterson (ed.), *Genomics of the Saccharinae*, Plant Genetics and Genomics: Crops and Models 11, DOI 10.1007/978-1-4419-5947-8_6,

1 Introduction

An individual's genome is its fixed blueprint, and this blueprint is slightly variable in DNA sequence across a population. The conversion of DNA-encoded information contained within nuclear and nonnuclear DNA, however, is not static. Throughout the genome, chromatin is dynamically relaxed allowing for the transcription to generate ribonucleic acid (RNA) molecules with several functions. For example, mRNA is incorporated into RNA (rRNA, tRNA)-protein complexes and translated into peptide sequences. RNA can even have a direct effect on the transcriptional silencing of genes (e.g., siRNA, miRNA). All expressed RNA molecules and their molecular counts are a cell's *transcriptome*. A genome position from which functional RNA (protein encoding or nonprotein encoding) is transcribed is one definition of a gene (Brosius 2009), which is not as easily defined as it has been in the past given the results of studies including the human ENCODE project (Birney et al. 2007; Rosenbloom et al. 2010). In this chapter we review gene activity from three Saccharinae taxa: sugarcane, sorghum, and *Miscanthus*.

Due to the limitations of current sequencing technology, it is often difficult to obtain a full length RNA transcript sequence from a purified RNA sample. Thus a partially sequenced RNA molecule is often what one finds when analyzing an expressed sequence tag (EST). In order to generate a set of ESTs, RNA is purified from a tissue of interest and converted to cDNA through reverse transcription and second strand synthesis. In classical Sanger based sequencing (Sanger et al. 1977), cDNA is cloned into a high-copy *E. coli* plasmid, amplified, and sequenced. More recent techniques (e.g., next-generation sequencing) include novel methods of cDNA amplification and sequencing and are highly parallel, allowing for the generation of many more but smaller ESTs than classical approaches with tag counts reaching into the millions (Mardis 2008). Sanger-based EST sequences can be trimmed for low-quality sequence, masked for *E. coli* and vector contamination, and assembled into nonredundant contigs (unigenes) using tools such as Phrap/cross_match (Ewing and Green 1998; Ewing et al. 1998), CAP3 (Huang and Madan 1999), and EGassembler (Masoudi-Nejad et al. 2006). Sequence assembly and analysis tools for next-generation sequence datasets are under intense development (see Miller et al. 2010 for review). EST datasets can be found in public repositories such as the NCBI EST and Short Read Archive (SRA) databases (http://www.ncbi.nlm.nih.gov/) as well as public repositories such as SUCEST (http://sucest-fun.org) and CSGR (http://csgr.agtec.uga.edu/). Below we discuss the known EST transcriptomes and summed gene space from sorghum and sugarcane studies.

The collection of unique EST assemblies from a given tissue is a snapshot of the tissue's transcriptome. The intersection of all transcriptomes is a measure of an organism's known gene space. Within the Saccharinae clade, the *Sorghum bicolor* (cv. BTx623) genome has been sequenced (Paterson et al. 2009) so the sorghum gene space has also been extended with gene predictions. Once gene space has been determined, it can be used as a baseline for comparing relative quantities of RNA between samples (transcriptional profiling). The direct counting of sequenced ESTs from an RNA sample mapped to the set of nonredundant genes is a form of transcriptional

profiling known as RNA-seq. Another form of transcriptional profiling uses the ordered attachment of DNA molecular probes to a glass surface (microarrays). By measuring relative hybridization intensities between complex mixtures of RNA molecules to specific probes of known sequence, it is possible to dissect differences in the transcriptome between samples. A discussion of sugarcane microarray studies will follow below.

2 Overview of EST Projects

2.1 Sorghum Gene Space and ESTs

A significant body of work was conducted on the sorghum transcriptome prior to its genome being sequenced in 2009. In April 2010, there were 209,828 *Sorghum bicolor*, 1,965 *Sorghum halepense* (Johnson grass), and 20,881 *Sorghum propinquum* sequences in the NCBI EST database. Approximately half (117,682) of the described ESTs were generated by a single group from multiple RNA sources including seedling tissue (light/dark grown), and plants subjected to drought stress, ovaries, embryos, and immature panicles (Pratt et al. 2005). Nine *S. bicolor* tissue specific EST libraries (OV1, WS1, DG1, PIC1, OV2, PI1, IP1, EM1, LG1), two *S. propinquum* libraries (FM1, RHIZ2), and one *S. halepense* library (RHIZ1) can be mined with estMiner at Comparative Saccharinae Genomics Resource (CSGR; http://csgr.agtec.uga.edu/estMiner/estMiner.jsp). The published version of the sorghum genome (Paterson et al. 2009) provides evidence for 34,496 gene models using many of these EST sequences, and homology-based as well as ab initio gene predictions. In addition, 149 known or predicted miRNA sites have been mapped to the sorghum genome (Paterson et al. 2009). The sorghum gene space is sufficiently mature for the development of additional microarray resources or the mapping of RNA-seq data.

Until other genome assemblies become available, the sorghum genome provides a temporary comparative baseline for members of Saccharinae clade. For example, while it is not possible to identify sorghum-specific transcripts in the sugarcane genome, it is possible to putatively identify sugarcane-specific transcripts by the simple fact that they do not map to the sorghum genome. Of course, there will be errors in this type of comparative analysis when genes rapidly diverge or are present in assembly gaps.

2.2 Sugarcane EST Datasets

The first study published on the sugarcane transcriptome was carried out by a group from South Africa (Carson and Botha 2000). The authors prepared a small collection of ESTs from leaf roll and stem from the cultivar Nco376. Although it was a

small library, it was sufficient to infer diverse physiological functions and attract an interest to EST sequencing projects. A Brazilian consortium developed the largest data set of sugarcane ESTs, named Sugarcane Expression Sequence Tag (SUCEST) (Vettore et al. 2001). A total of 259,325 clones were generated and sequenced from their 5′ end and 32,364 also had their 3′ end region sequenced, in sum yielding 237,954 ESTs of good quality. These ESTs were assembled into 43,141 EST clusters. This collection was produced from 26 libraries from the cultivars SP70-1143, SP80-3280, SP80-87432, PB5211 × P57150-4, CB47-89, RB855205, RB845298, and RB805028, sampling from several different tissues, such as apical meristem, lateral bud, internodes from immature plants, seeds, flowers, leaves, roots, and roots infected with endophytic nitrogen fixing bacteria. An analysis of the 43,141 clusters, named Sugarcane Assembled Sequence (SAS), indicates that the SUCEST project may have identified 33,620 expressed sugarcane genes. Since the sugarcane genome is estimated to contain 35,000 genes, SUCEST data may represent >90 % of sugarcane expressed genes (Vettore et al. 2003). These SAS were compared with sequences present in public databases and categorized with respect to functionally annotated genes in other organisms (Vettore et al. 2003). The sequences were grouped into eighteen broad categories of biological roles. Almost 50 % of total SAS were classified in five major categories: cellular dynamics, stress response, protein metabolism, bioenergetics, and signal transduction. However, approximately 25 % of total SAS were classified as putative proteins or unable to classify. These SAS may represent new genes, in which they could represent sugarcane- and/or grass-specific genes but also noncoding regulatory RNAs. SUCEST was the starting point to the SUCAST project (Sugarcane Signal Transduction), which focused in the identification of signal transduction related genes. The SUCAST catalogue contains around 3,500 SAS organized in 29 categories and 409 subcategories (Papini-Terzi et al. 2005). The SAS present in this catalogue can be divided in functional categories, such as: development, cell cycle, stress and pathogenicity.

Another important sugarcane EST data set was generated by Australian researchers (Casu et al. 2004; Casu et al. 2003). They initially prepared cDNA libraries from sucrose accumulating maturing stem and immature stem of sugarcane cultivar Q117. A total of 7,242 ESTs were generated from the former library and a smaller sample, 1,082 ESTs, was obtained from the latter. A BLAST search of the total ESTs obtained retrieved 3,845 ESTs with no significant match to known sequence or matches with unknown function sequences. Besides, only about 2 % of the ESTs were carbohydrate metabolism related sequences, indicating that these transcripts are not abundant even in sucrose accumulating stem internodes. Within these carbohydrate metabolism related sequences, sugar transport proteins were highly abundant in the maturing stem. The same research group obtained a set of 747 ESTs from roots (cultivar Q117) treated with the defense-regulator methyl jasmonate (MJ) (Bower et al. 2005). A highly significant portion of these ESTs could not be classified in a functional category. Primary metabolism and protein metabolism were the two functional categories most represented in this EST collection, and only a small portion has putative function related to plant defense and stress. In combination with microarray and northern blot experiments these authors were able to identify ESTs induced by MJ treatment.

A group in the United States carried out an EST sequence project from apices, leaves and seventh internode from cultivar CP72-2086 (Ma et al. 2004). A total of 9,216 clones were sequenced resulting in 3,401 nonredundant ESTs. The ESTs were classified into 12 categories: cellular structure, signal transduction, DNA metabolism, RNA expression, protein expression, stress/defense, metabolism, growth, transport, others, unknown, and no hit. The most abundant category was "unknown," followed by protein expression. A portion of the collection could not be classified and may represent new sequences.

A recent EST project from an Indian group was carried out aiming to identify ESTs expressed in response to water deficit in leaves from the cultivar CoS 767 and red-rot infection from stem in the cultivar Co 1148 (Gupta et al. 2010). The group also prepared a general cDNA library from root, stem, leaf whorl, and mature leaf from field-grown plants of cultivar CoS 767. From the CoS 767 library, 25,382 good quality sequences were obtained. After assembly with other ESTs collections, 4,002 new ESTs were reported. The sequencing, analysis and assembling of 1,440 ESTs from the red-rot library retrieved 102 clusters, in which 85 is thought to be preferentially expressed in red-rot infection.

The current Sugarcane Gene Index (3.0, April 2010 http://compbio.dfci.harvard.edu/cgi-bin/tgi/gimain.pl?gudb=s_officinarum) has clustered 282,683 ESTs and 499 expressed transcripts (fully sequenced cDNAs), resulting in 42,377 theoretical contigs, 78,924 singleton ESTs, and 41 singleton expressed transcripts.

2.3 Other Saccharinae Transcriptome Datasets

While much is known about sorghum and sugarcane gene space, very little work has been applied to other members of the Saccharinae clade. In April 2010, the only significant additional dataset were three Illumina sequencing runs of small RNA libraries from *Miscanthus* × *giganteu*s rhizomes (SRX016338; five million reads; 169.5 Mbp), flowers (SRX016337; four million reads; 133.1 Mbp), and leaves (SRX016336; four million reads; 137 Mbp). This study may point to a major role that new sequencing technologies will play in orphan genomes such as *Miscanthus* × *giganteu*s.

3 Comparative EST Analysis of Sorghum and Sugarcane

Comparative genome and transcriptome data analysis has been critical for a better understanding of genome structure, evolution and gene function. The *Sorghum bicolor* genome has a smaller physical size than those of many grasses and it is diploid, with lower redundancy than polyploid genomes including sugarcane and *Miscanthus* (Swaminathan et al. 2010). The *Miscanthus*×*giganteus* genome has been surveyed (84 Mbp in 366,448 454 FLX reads), but sorghum provides one of the best functional genomics models for the Saccharinae and other C4 grasses

Table 6.1 Genome, genes and EST sequences repositories used in this study

ESTs	
Saccharinae	http://sucest-fun.org
	http://www.ncbi.nlm.nih.gov/nucest
Genomes and genes	
Sorghum bicolor	http://genome.jgi-psf.org/Sorbi1/Sorbi1.home.html
Oryza sativa	http://rapdb.dna.affrc.go.jp/
Zea mays	http://ftp.maizesequence.org/
Brachypodium distachyon	http://www.brachypodium.org/
Miscanthus spp.	http://genomebiology.com/2010/11/2/R12

Table 6.2 Expressed sequence tags corresponding to *Saccharum* and *Sorghum* taxa in GenBank as of April 7, 2010

	# sequences	Average size
Saccharum ESTs in GenBank	283,157	612
Saccharum ESTs in GenBank from SUCEST Project	238,208	641
Saccharum ESTs in GenBank from others	44,954	456
Sorghum ESTs in GenBank	209,828	551

(Swaminathan et al. 2010). In a comparison of large genomic fragments cloned in Bacterial Artificial Chromosomes (BACs) of the grasses maize, rice, sorghum and sugarcane (Jannoo et al. 2007) *Sorghum* and sugarcane sequence alignments revealed a high general colinearity and structure conservation, although some genes were found only in sorghum and a few rearrangements were identified. *Sorghum* and sugarcane presented considerable homology in the noncoding regions and sugarcane contains larger transposable elements (Jannoo et al. 2007).

GenBank, DDBJ, and EMBL are the main repositories for genome, transcriptome, and proteome sequences and are synchronized periodically. For interspecies comparisons, genomes or EST sequences were downloaded from the Web sites listed in Table 6.1, including the sorghum coding sequences (CDS).

To obtain the number of specific ESTs for *Saccharum* species for comparison with other species, the complete set of *Saccharum* EST sequences (283,157 ESTs) deposited in GenBank was compared against the available *Sorghum* EST sequences (209,828 ESTs) (Table 6.2).

Saccharum and *Sorghum* ESTs were compared to *Zea mays*, *Oryza sativa*, *Miscanthus*, and *Brachypodium distachyon* ESTs. The analysis was performed using MEGABLAST at a stringency of 1e-25 and WordSize equal to 25. Table 6.3 summarizes the number of hits and no hits for each Genome or CDS set, including the percentage of coverage and identity average. The Table also indicates the total number of common and specific ESTs from *Saccharum* in relation to the other grasses. A total of 205,061 *Saccharum* ESTs found a match in the sorghum genome, and 200,207 *Saccharum* ESTs found a match in the sorghum CDS. From 78,000 to 83,000 *Saccharum* ESTs found “no match” against the sorghum sequences at this stringency and Wordsize. A search of these so-called “no match” sequences against

Table 6.3 *Saccharum* ESTs were compared with available ESTs from other grasses (*Sorghum*, maize, *Oryza sativa*, *Brachypodium* and *Miscanthus*) and *Sorghum* CDS to define ESTs specific to *Saccharum* species

Match—Saccharum ESTs	Hits	No hits	Average coverage %	Average identity %
(1) Blastn *Saccharum* ESTs vs. *Sorghum* Genome	205,061	78,096	72	92
(2) Blastn *Saccharum* ESTs vs. *Sorghum* CDS	200,207	82,950	80	94
"No match"—*Saccharum* ESTs in *Sorghum* Genome				
(3) Blastn *Saccharum* ESTsNoMatch vs. Maize Genome	7,761	70,335	42	93
(4) Blastn *Saccharum* ESTsNoMatch vs. *Oryza sativa* jp. Genome	831	77,265	31	93
(5) Blastn *Saccharum* ESTsNoMatch vs. *Brachypodium* Genome	9,184	68,912	30	90
(6) Blastn *Saccharum* ESTsNoMatch vs. *Miscanthus* Genome (1 %)	13,690	64,406	30	93
(7) Total *Saccharum* ESTs Match against the genomes	225,967 (79.8 %)			
(8) Total *Saccharum* EST No Match against the genomes above	57,190 (21.2 %)			

Table 6.4 Alignment of SAS (Sugarcane Assembled Sequences) to *Sorghum* chromosomes and supercontigs

Description	Total SAS	Unique SAS	Avg. coverage of SAS (bp)
SAS aligned to sorghum genome	30,619	25,132	802
SAS not aligned to sorghum genome.	18,009		

The alignments were performed using Sim4 at default parameters. The alignments with a coverage of SAS less than 80 % where filtered out. The column "Unique SAS" summarizes the number of distinct SAS mapped to the sorghum genome

the maize, *Oryza*, *Brachypodium*, and *Miscanthus* genomes still yielded 57,190 Saccharum "no-match" or Saccharum specific sequences. *Miscanthus* was the genome that had largest number of matches to otherwise *Saccharum* specific ESTs, consistent with the more recent common ancestry of *Saccharum–Miscanthus* than the remaining species.

The SUCEST SAS database and the JGI *Sorghum* genome data were used to obtain the number of *Saccharum* transcripts that mapped to the sorghum genome (Table 6.4). The SAS dataset is clusters of ESTs assembled by the SUCEST project (Vettore et al. 2001) and contains 43,141 clusters. The *Sorghum bicolor* genome comprises ten chromosomes and 3,294 supercontigs. While there are fewer SAS per chromosome than sorghum genes, the number of SAS mapped per chromosome parallels the distribution of number of sorghum genes per chromosome (Fig. 6.1). This is an interesting feature that reinforces the observed conservation and similarity of these two species.

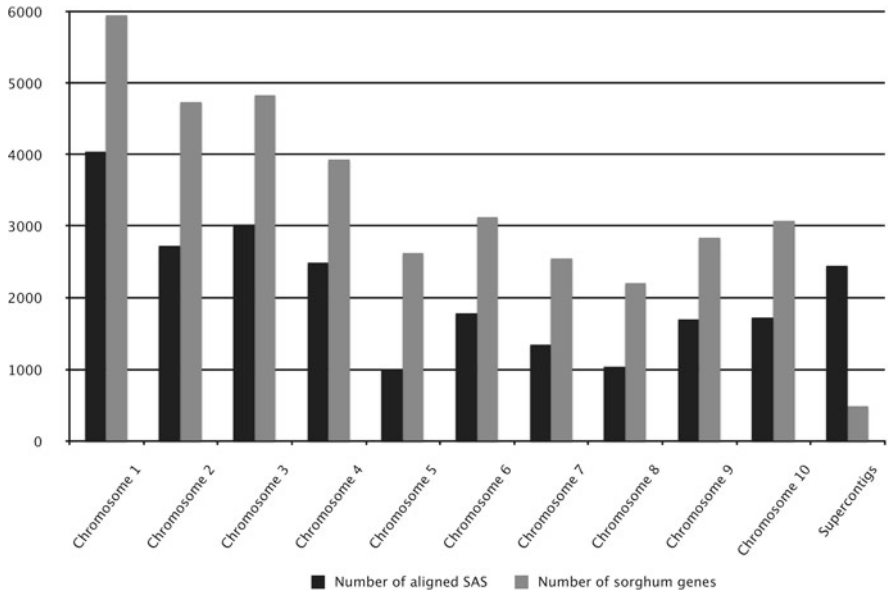

Fig. 6.1 Distribution of SAS along the sorghum chromosomes. The SAS were aligned using Sim4 at $W = 12$ and a cutoff of coverage for the SAS = 90 %. The number of mapped SAS is lower than the number of *Sorghum* genes. However, they follow the distribution of *Sorghum* genes per chromosomes

A total of 38,967 sugarcane proteins were predicted using ESTScan (Iseli et al. 1999) with the *Zea mays* matrix. To estimate the number of putative orthologs to the predicted 36,338 sorghum proteins, InParanoid was used at default parameters (Remm et al. 2001) (Table 6.5). The predicted number of Saccharum proteins with homologs in Sorghum was estimated to be 16,444 and the predicted number of sorghum proteins with homologs in Saccharum was 25,363, yielding a total number of orthologous groups of 10,432. A possible explanation of the lower number of Saccharum proteins with homologs in sorghum may be that SAS currently may contain copies of similar genes or that some SAS are incomplete sequences that map to the same sorghum gene.

4 Microarray Data

4.1 Saccharum Microarray Studies

The sugarcane transcriptome has been extensively analyzed using nylon arrays, cDNA microarrays and oligonucleotide arrays. Using nylon filter microarrays Nogueira et al. (2003), analyzed the expression profile of 1,536 sugarcane ESTs in

Table 6.5 Prediction of *Saccharum* proteins and orthologs to sorghum

Description	Total	Average size
Proteins in *Sorghum* bicolor	36,338	202
Predicted proteins in *Saccharum*	38,967	387
Predicted *Saccharum* proteins with homologs in *Sorghum*	16,444	
Sorghum proteins with homologs in *Saccharum*	25,363	
Total groups of orthologs	10,432	

samples exposed to cold for up to 48 h. They identified 59 high quality ESTs showing significantly altered expression during cold treatment. Thirty-four were up regulated, including five polyubiquitin proteins (CA097514.1, CA101228.1, CA064748.1, BU103666, and CA0974371.1) , a cellulose synthase-4 (CA097873.1), a xanthine dehydrogenase (XDH) (CA102747.1), a copper/zinc SOD (CA120040.1), a pyruvate orthophosphate dikinase (PPDK) (CA097241.1), and NADP-dependent malic enzyme (NADP-ME) (CA097270.1), indicating possible maintenance of photosynthetic processes during cold stress (Nogueira et al. 2003).

Environmental and developmental plant responses depend on intricate network signals of different cellular components. To compile a signal transduction database for sugarcane, 3,563 SAS involved in several aspects of signal transduction, transcription, development, cell cycle, stress responses, and pathogen interaction were categorized into the SUCAST Catalogue (http://sucest-fun.org). The ESTs were grouped into 29 main categories subdivided into 409 subcategories (Papini-Terzi et al. 2005; Souza et al. 2001).

To explore individual variation in sugarcane field-grown plants, cDNA microarrays containing 1,280 SAS from the SUCAST catalogue were used to define tissue expression profiles and transcript abundance in sugarcane organs (flowers, roots, leaves, lateral buds, and first and fourth internodes) (Papini-Terzi et al. 2005). Differential expression was more pronounced in leaves (8.84 %), roots (11.78 %) and fourth internodes (8.26 %). Flower samples showed four transcription factors to be highly expressed: a GARP transfactor (CA162928.1), an AP2 (CA116150.1), a zinc finger (CA110100.1) and a MADS-box domain-containing proteins (CA172499.1) that play important roles in organ development were particularly pronounced. Three receptor genes (CA116933.1, CA124905.1, CA070313.1) were preferentially expressed in leaves (receptor-like kinase) and components of biosynthetic pathways of salicylic acid (phenylalanine ammonia-lyase; CA132523.1), abscisic acid (zeaxanthin epoxidase; CA125475.1), and ethylene ACC oxidase (CA130173.1), ACC synthase (CA099229.1). Four nitrilases (CA125195.1, CA095693.1, CA103229.1, and CA132900.1) and three lipoxygenases (CA129965.1, CA130399.1, and CA133468.1) showed prevalent expression in root tissues. In the fourth internode a gene of 3-*O*-methyltransferase (COMT) (CA125200.1) was expressed which is related primarily to lignin biosynthesis.

The cDNA microarray approach described above was also used to monitor expression changes of 1,545 genes in plantlets submitted to biotic and abiotic stress (Rocha et al. 2007). Using the outlier method HTSelf (Vêncio and Koide 2005),

SOM (Self-Organizing Maps) and qRT-PCR, the authors found 485, 146, 158 and 68 differentially expressed genes in sugarcane plants submitted to drought, phosphate starvation, ABA and MeJA respectively. Forty-four percent of the differentially expressed genes were drought responsive. Signaling pathways related to cold and drought have a high degree of overlap. For example, two SAS encoding low temperature induced (LTI) proteins (CA186860 and CA096029) were upregulated in response to water deprivation. Other overlaps were found between ABA and drought signaling. Among the genes induced were two delta-12 oleate desaturase (CA189695 and CA179715), one *S*-adenosylmethionine decarboxylase (CA189868) and a PP2C-like protein phosphatase (CA147516) homologous to ABI1 and ABI2 ABA-responsive genes of *Arabidopsis*. ABA treatment induced a glycine-rich protein (CA189990) and two GTPases (CA095849 and CA134244) similar to Rab11 of *Arabidopsis*. In the MeJA treatment, two H4 (CA127099 and CA106117) histones and one H2B (CA120037) were upregulated indicating a possibly regulation by chromatin remodeling. Phosphate starvation altered the expression the majority of genes after 6 h of treatment.

Increasing sucrose in sugarcane culms is one of the main targets of breeding programs. To identify genes associated with sucrose content Papini-Terzi and colleagues (2009) evaluated the transcriptome of genotypes of crosses between *S. officinarum* and *S. spontaneum* and commercial varieties SP80-180 and SP80-4966. Individuals with extreme values of sugar content were compared. Among the genes differentially expressed were those related to hormone signaling (auxin, ethylene, jasmonates), stress responses (drought, cold, oxidative), cell wall metabolism, calcium metabolism, protein kinases, protein phosphatases, and transcription factors.

Fifty-four genes represented by protein kinases (PKs), protein phosphatases (PPases) or receptor-like kinases (RLKs) were differentially expressed in high Brix plants and during culm development. PKs from the SNF1/SnRK family were highly represented among the differentially expressed genes. This family regulates the expression of genes related to carbohydrate metabolism in yeast and in plants (Halford and Hey 2009). In mature internodes where sucrose accumulation is higher some members of this family were found to be expressed at a lower level. Additionally, four 14-3-3 (CA119664.1, BU103695.1, CA133114.1, and CA132593.1) proteins of the GF14 type also were found altered. These proteins participate in the signaling mediated by SnRK (Sugden et al. 1999; Torosera et al. 1998). A large overlap of gene expression changes were observed when sucrose content and drought response datasets were compared. Phosphorylation events were also implicated to have a role in drought responses. Sixty-nine genes associated with sucrose content were regulated by drought and eleven by ABA. One protein phosphatase of the PP2C family (CA147516.1) was less expressed in high Brix, reduced in the mature internodes and induced by drought and ABA. A member of the CBL interacting protein kinase (*ScCIPK-21*) gene family (CA142458.1) found to be more expressed in high Brix and mature internodes was much more induced during culm development in high Brix plants. This may be indicative that the induction of this gene leads to higher sucrose levels.

Several transcription factors (TFs) were found to be differentially regulated. In immature internodes, two members of the homeobox knotted 1-homeodomain (CA116948.1 and CA078708.1), a TF related to ARF6 (auxin response factor 6a) (CA119432.1) and three auxin response proteins (CA146855.1, scvplr2005h03.g, CA151389.1.g, CA122611.1.g) were more expressed than in mature internodes. A second ARF6 (CA113224.1), a NAM (no apical meristem) (CA127180.1) and an EIL (ethylene insensitive3-like) (CA147144.1) showed differentially expressed when high and low brix genotypes were compared.

Cell wall metabolism genes have also been evaluated at the expression level in sugarcane. Expansins may act in the relaxation of the cell wall of plants and in sugarcane two of them were found to be differentially expressed when high and low brix plants were compared. A xyloglucan endo-β-1, 4 glucanase (XTH) (CA117385.1) and three genes (CA119495.1, CA132523.1, and CA079959.1) of the lignin biosynthesis pathway were associated with sucrose content. These findings indicate that cell wall metabolism may be altered in plants improved for sucrose content (Papini-Terzi et al. 2009).

Using plants from the same crossing between commercial varieties described above (SP80-180 vs. SP80-4966), Felix and colleagues (2009) profiled mature leaves. Among the genes altered, the group found an omega-3 fatty acid desaturase (FAD8) (CA079174), a putative receptor-like serine/threonine kinase (ScBAK1) (CA156919), and a Myb domain transcription factor LHY/CCA1 (CA190110). Transcripts enriched in mature leaves of low sugar content plants included three 14-3-3 like proteins (CA146811, CA132593, and CA133114), two proteins involved in inositol metabolism [O-methyltransferase (CA125200) and Inositol-4,5-trisphosphate phosphatase (CA185029)], a SNF1-related protein (SnRK1) (CA279976), eight stress-related proteins (CA122463, CA160294, CA186860, CA239336, CA119392, CA124270, CA135201, and CA127342), two transcription factors (CA110838 and CA093881), and an F-box TIR-1 (CA096709). All data described above can be accessed through the SUCEST-FUN Database (http://sucest-fun.org) (Felix et al. 2009).

Other sugarcane microarray platforms not included in the SUCEST-FUN Project can be found in the public repository GEO-NCBI (Table 6.6). An Affymetrix Genechip for the Sugarcane Genome containing 8,236 probe sets was created that can be used to monitor gene expression for approximately 6,024 distinct genes. Using this platform Casu and colleagues (2007) identified over 100 differentially expressed transcripts during development. They also showed that cellulose synthase (CesA) and cellulose synthase-like (Csl) gene families were coordinately expressed during sugarcane stem development (Casu et al. 2007).

Genes responsive to sucrose and hexose accumulation in sugarcane leaves that had their sugar concentration altered by excision and cold-girdling were identified. Sugar accumulation was correlated to a decrease in photosynthesis. The authors identified 74 differentially expressed genes involved in primary and secondary metabolic pathways and 21 genes associated with photosystems I and II. Increased expression of sugar transporters (BU925715 and AY165599) and a MAPK (CA272048) and the decreased expression of two SnRK1 regulatory subunit beta-1

Table 6.6 Summary of the microarray data available for sugarcane

Platform[a]	Record GEO NCBI[b]	Features[c]	Samples[d]	Series GEO NCBI[e]	Treatment[f]	Variety[g]	Experiment description[h]	References[i]
Sugarcane								
Sugarcane ESTs nylon arrays I	GPL210	1,536	12	*GSE6442*	Cold	SP80-3280	Plantlets samples exposed to cold for up to 48 h	Nogueira et al. (2003)
SUCAST Sugarcane 2208 v1.0 (cDNA Microarray)	GPL3799	2,208	16	*GSE4966*	Phosphate starvation	SP80-3280	Roots from six plants from each treatment (0 and 250 μM Pi) were harvested 6, 12, 24 and 48 h after exposure to phosphate starvation	Rocha et al. (2007)
			8	*GSE4967*	Response to herbivory by *Diatraea saccharalis*	SP80-3280	Leaves were collected after 30 min and 24 h of exposure to herbivory by *Diatraea saccharalis* for the control and experimental groups	
			16	*GSE4968*	ABA treatment	SP80-3280	ABA was added to the culture medium to a final concentration of a 100 μmol L^{-1}	
			12	*GSE4969*	MeJa treatment	SP80-3280	Plantlets were then sprayed with a 100 μmol L^{-1} MeJA solution, whereas control plantlets were treated with distilled water. Leaves were collected 0, 1, 6and 12 h after exposure to MeJA	
			8	*GSE4970*	Response to N_2-fixing endophytic bacteria association	SP70-1143	In vitro-grown SP70-1143 rooted sugarcane plantlets were inoculated as described by James et al. (1994) with 0.1 mL of 106–107 bacterial suspension	
			12	*GSE4971*	Drought response	SP90-1638	Leaves were collected 24, 72 and 120 h after exposure to drought conditions for the control and experimental groups	

SUCAST 1920v3 [*Saccharum officinarum*] (cDNA Microarray)	GPL1376	1,920	26	*GSE1702*	Transcripts abundance in sugarcane tissues	SP80-3280	The abundance of transcripts among six different sugarcane tissues (flowers, roots, leaves, lateral buds, first and fourth internodes)	Papini-Terzi et al. (2005)
			4	*GSE4233*	Signal Transduction Components in a Sugarcane Population Segregated for Sugar Content	SP80-3280	Leaves from sugarcane population segregated in relation to high and low brix	
Affymetrix Sugarcane Genome Array	GPL3844	6,024	12	*GSE5021*	Transcripts associated with cell wall metabolism and development in the sugarcane stem	Q117 grown field	Leaves of sugarcane in response to sugar accumulation in meristem, internodes 1–3; internode 8; internode 20	Casu et al. (2007)
			8	*GSE11934*	Leaves of sugarcane in response to sugar accumulation	*Saccharum* spp. (L.) hybrid cv. N19	Cold-girdles were attached to sugarcane leaves ($n=4$) for a period of 56 h prior to harvest	McCormick et al. (2008)

All published sugarcane and sorghum microarrays were catalogued based on MIAME compliant (Brazma et al. 2001) public repositories such as Gene Expression Omnibus (GEO-NCBI), Center for Information Biology Gene Expression Database (CIBEX), Microarray Gene Expression Data Society (MGED), ArrayExpress and the Lee H. Pratt Web site

[a]Array type
[b]Record available in GEO NCBI
[c]Number features on array type
[d]Number samples available on platform
[e]Number series sample in GEO NCBI
[f]Brief description on treatment
[g]Variety used in experiment
[h]Brief experiment description
[i]Reference of data

genes (CF577254 and CA093131) indicate that sugar accumulation induced a signaling cascade. Four genes related to trehalose metabolism including a trehalose 6-phosphate synthase (CA176198 and CF573831) and a trehalose 6-phosphate phosphatase (CA086449 and CF575026) pointed to these genes as potential targets to disable sugar signaling mechanisms that might inhibit sugar accumulation (McCormick et al. 2008).

4.2 Sorghum Microarray Studies

Gene expression studies have been reported that describe sorghum responses to herbivory, biotic and abiotic stress, which can be accessed through a MIAME compliant database (Brazma et al. 2001) (Table 6.7).

To understand if attack by a phloem-feeding greenbug aphid activates jasmonic acid (JA) and salicylic acid (SA) regulated genes, a collection of 672 cDNAs was obtained using a subtraction method. A total of 82 responsive transcripts were found that encode proteins involved in direct defense, defense signaling, oxidative burst, secondary metabolism, abiotic stress, cell maintenance, and photosynthesis (Zhu-Salzman et al. 2004). The transcriptome profile of the highly greenbug-resistant *Sorghum* M627 line was also evaluated. In this case, 157 differentially expressed transcripts were found that belong to nine functional categories (Sung-Jing et al. 2006).

In relation to abiotic and biotic stress, Buchanan et al. (2005) exposed seedlings of *Sorghum bicolor* to high salinity (150 mM NaCl), osmotic stress (20 % polyethylene glycol) and ABA. Gene expression was examined in roots and shoots at 3 and 27 h post treatment. A sorghum cDNA microarray with 12,982 unique gene clusters was used and 2,200 genes were found altered in response to the treatments (Buchanan et al. 2005).

Using the same microarray system Salzman et al. (2005) profiled the response to salicylic acid (SA), methyl jasmonate (MeJA) and the ethylene precursor aminocyclopropane carboxylic acid. *Sorghum* bicolor L. Moench cv BTx623 seedling root and shoot tissues were evaluated at 3 and 27 h after treatment. Data from over one hundred hybridizations and quantitative PCR analysis for 171 genes showed both an antagonism and synergism by SA and MeJA. The antagonism effect means that the great majority of the 18 genes MeJA induced, showed lower induction when MeJA + SA were applied. The synergistic effect was showing by 22 genes (greater than the sum of individual inductions) to SA + MeJA treatment than by SA and MeJA alone.

For other grasses such as *Miscanthus*, there is no transcriptome profile information available at this time. Comparisons of this grass against sugarcane will be very informative, since *Miscanthus* is expected to be very similar at the DNA sequence level to sugarcane.

Recently, the sorghum genome sequence has been obtained (Paterson et al. 2009). Several groups have started to sequence sugarcane using next-generation technologies and some *Miscanthus* shot-gun sequences have also been recently

Table 6.7 Summary of the microarray data available for *Sorghum*

Platform[a]	Record[b]	Features[c]	Expressed genes[d]	Treatment[e]	Variety[f]	Experiment description[g]	Reference[h]
Sorghum							
cDNA microarray	http://fungen.org/*Sorghum*.htm	12,982	2,200	High salinity, osmotic stress, and ABA	Seedlings of *Sorghum* bicolor L. Moench cv BTx623	3 and 27 h treatment high salinity (150 mM NaCl), osmotic stress (20 % polyethylene glycol)	Buchanan et al. (2005)
		12,982	6,438	salicylic acid (SA), methyl jasmonate (MeJA), and the ethylene precursor aminocyclopro-pane	Seedlings of *Sorghum* bicolor L. Moench cv BTx623	3 and 27 h treatment salicylic acid (SA), methyl jasmonate (MeJA), and the ethylene precursor aminocyclopropane	Salzman et al. (2005)
cDNA glass slide	Not available	672	128	Herbivory with *Schizaphis graminum*	Sseedlings of *Sorghum* bicolor ATx399_RTx430	48 h after phloem-feeding greenbug aphid introduction	Zhu-Salzman et al. (2004)
		3,508	157	Herbivory with *Schizaphis graminum*	Sseedlings of *Sorghum* bicolor M627 and Tx7000	12, 24, and 72 h after phloem-feeding greenbug aphid introduction	Park et al. (2006)

All published sugarcane and sorghum microarrays were catalogued based on MIAME compliant (Brazma et al. 2001) public repositories such as Gene Expression Omnibus (GEO-NCBI), Center for Information Biology Gene Expression Database (CIBEX), Microarray Gene Expression Data Society (MGED), ArrayExpress and the Lee H. Pratt Web site

[a]Array type
[b]Record available
[c]Number features on array type
[d]Number genes differentially expressed
[e]Brief description on treatment
[f]Variety used in experiment
[g]Brief experiment description
[h]Reference of data

released. To compare transcriptome data it will be important to conduct a careful identification of putative orthologs. The microarray data provided so far together with the knowledge of promoter sequences will allow for a better understanding of regulatory pathways and gene networks related to agronomic traits of interest. This will be an important step in defining gene targets to produce new varieties with improved sucrose content, high biomass and increased energy content for biofuel production.

5 Final Considerations

The Brazilian consortium developed the main dataset of sugarcane ESTs called SUCEST representing more than 90 % of sugarcane expressed genes. Due to this ESTs collection and smaller ones, a lot of information about transcriptome profile of sugarcane in different development stages and submitted to several stress conditions are available and some signaling pathways have shown a high degree of overlap. Among these, the phosphorylation and dephosphorylation events can be found related to sucrose content, drought response, and ABA. Some transcription factors and expansins can be differentially regulated during culm development and among contrasting sucrose content cultivars. Genes related to trehalose metabolism were regulated during sugar accumulation and cellulose synthase and cellulose synthase-like genes families were upregulated during sugarcane stem development.

The transcriptome profile of Saccharinae clade is very important to understand how different conditions of plant development and stress can affect agronomic traits of interest. However, to compare transcriptome data it is necessary to conduct a careful identification of alleles and putative orthologs.

For the future, the challenge will be to obtain the sugarcane BAC library and shotgun sequences through the Next-generation sequencing, and thus, the microarray data provided so far, together with the knowledge of promoter sequences, will allow the better understanding of regulatory pathways and gene networks. This will be an important step in defining gene targets to produce news varieties with improved sucrose content, high biomass, and increased energy content for biofuel production.

References

Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, Kuehn MS, Taylor CM, Neph S, Koch CM, Asthana S, Malhotra A, Adzhubei I, Greenbaum JA, Andrews RM, Flicek P, Boyle PJ, Cao H, Carter NP, Clelland GK, Davis S, Day N, Dhami P, Dillon SC, Dorschner MO, Fiegler H, Giresi PG, Goldy J, Hawrylycz M, Haydock A, Humbert R, James KD, Johnson BE, Johnson EM, Frum TT, Rosenzweig ER, Karnani N, Lee K, Lefebvre GC, Navas PA, Neri F, Parker SC, Sabo PJ, Sandstrom R, Shafer A, Vetrie D, Weaver M, Wilcox S, Yu M, Collins FS, Dekker J, Lieb JD, Tullius TD, Crawford GE, Sunyaev S, Noble WS, Dunham I, Denoeud F, Reymond A,

Kapranov P, Rozowsky J, Zheng D, Castelo R, Frankish A, Harrow J, Ghosh S, Sandelin A, Hofacker IL, Baertsch R, Keefe D, Dike S, Cheng J, Hirsch HA, Sekinger EA, Lagarde J, Abril JF, Shahab A, Flamm C, Fried C, Hackermuller J, Hertel J, Lindemeyer M, Missal K, Tanzer A, Washietl S, Korbel J, Emanuelsson O, Pedersen JS, Holroyd N, Taylor R, Swarbreck D, Matthews N, Dickson MC, Thomas DJ, Weirauch MT, Gilbert J, Drenkow J, Bell I, Zhao X, Srinivasan KG, Sung WK, Ooi HS, Chiu KP, Foissac S, Alioto T, Brent M, Pachter L, Tress ML, Valencia A, Choo SW, Choo CY, Ucla C, Manzano C, Wyss C, Cheung E, Clark TG, Brown JB, Ganesh M, Patel S, Tammana H, Chrast J, Henrichsen CN, Kai C, Kawai J, Nagalakshmi U, Wu J, Lian Z, Lian J, Newburger P, Zhang X, Bickel P, Mattick JS, Carninci P, Hayashizaki Y, Weissman S, Hubbard T, Myers RM, Rogers J, Stadler PF, Lowe TM, Wei CL, Ruan Y, Struhl K, Gerstein M, Antonarakis SE, Fu Y, Green ED, Karaoz U, Siepel A, Taylor J, Liefer LA, Wetterstrand KA, Good PJ, Feingold EA, Guyer MS, Cooper GM, Asimenos G, Dewey CN, Hou M, Nikolaev S, Montoya-Burgos JI, Loytynoja A, Whelan S, Pardi F, Massingham T, Huang H, Zhang NR, Holmes I, Mullikin JC, Ureta-Vidal A, Paten B, Seringhaus M, Church D, Rosenbloom K, Kent WJ, Stone EA, Batzoglou S, Goldman N, Hardison RC, Haussler D, Miller W, Sidow A, Trinklein ND, Zhang ZD, Barrera L, Stuart R, King DC, Ameur A, Enroth S, Bieda MC, Kim J, Bhinge AA, Jiang N, Liu J, Yao F, Vega VB, Lee CW, Ng P, Yang A, Moqtaderi Z, Zhu Z, Xu X, Squazzo S, Oberley MJ, Inman D, Singer MA, Richmond TA, Munn KJ, Rada-Iglesias A, Wallerman O, Komorowski J, Fowler JC, Couttet P, Bruce AW, Dovey OM, Ellis PD, Langford CF, Nix DA, Euskirchen G, Hartman S, Urban AE, Kraus P, Van Calcar S, Heintzman N, Kim TH, Wang K, Qu C, Hon G, Luna R, Glass CK, Rosenfeld MG, Aldred SF, Cooper SJ, Halees A, Lin JM, Shulha HP, Xu M, Haidar JN, Yu Y, Iyer VR, Green RD, Wadelius C, Farnham PJ, Ren B, Harte RA, Hinrichs AS, Trumbower H, Clawson H, Hillman-Jackson J, Zweig AS, Smith K, Thakkapallayil A, Barber G, Kuhn RM, Karolchik D, Armengol L, Bird CP, de Bakker PI, Kern AD, Lopez-Bigas N, Martin JD, Stranger BE, Woodroffe A, Davydov E, Dimas A, Eyras E, Hallgrimsdottir IB, Huppert J, Zody MC, Abecasis GR, Estivill X, Bouffard GG, Guan X, Hansen NF, Idol JR, Maduro VV, Maskeri B, McDowell JC, Park M, Thomas PJ, Young AC, Blakesley RW, Muzny DM, Sodergren E, Wheeler DA, Worley KC, Jiang H, Weinstock GM, Gibbs RA, Graves T, Fulton R, Mardis ER, Wilson RK, Clamp M, Cuff J, Gnerre S, Jaffe DB, Chang JL, Lindblad-Toh K, Lander ES, Koriabine M, Nefedov M, Osoegawa K, Yoshinaga Y, Zhu B, de Jong PJ (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447:799–816

Bower NI, Casu RE, Maclean DJ, Reverter A, Chapman SC (2005) Transcriptional response of sugarcane roots to methyl jasmonate. Plant Sci 168:761–772

Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FCP, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M (2001) Minimum information about a microarray experiment (MIAME) – toward standards for microarray data. Nat Genet 29:365–371

Brosius J (2009) The fragmented gene. Ann N Y Acad Sci 1178:186–193

Buchanan CD, Lim S, Salzman RA, Kagiampakis I, Morishige DT, Weers BD, Klein RR, Pratt LH, Cordonnier-Pratt M-M, Klein PE, Mullet JE (2005) Sorghum bicolor's transcriptome response to dehydration, high salinity and ABA. Plant Mol Biol 58:699–720

Carson D, Botha F (2000) Preliminary analysis of expressed sequence tags for sugarcane. Crop Sci 40:1769–1779

Casu RE, Dimmock CM, Chapman SC, Grof CP, McIntyre CL, Bonnett GD, Manners JM (2004) Identification of differentially expressed transcripts from maturing stem of sugarcane by in silico analysis of stem expressed sequence tags and gene expression profiling. Plant Mol Biol 54:503–517

Casu RE, Grof CP, Rae AL, McIntyre CL, Dimmock CM, Manners JM (2003) Identification of a novel sugar transporter homologue strongly expressed in maturing stem vascular tissues of sugarcane by expressed sequence tag and microarray analysis. Plant Mol Biol 52:371–386

Casu RE, Jarmey JM, Bonnett GD, Manners JM (2007) Identification of transcripts associated with cell wall metabolism and development in the stem of sugarcane by Affymetrix GeneChip Sugarcane Genome Array expression Profiling. Funct Integr Genomics 7:153–167

Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res 8:186–194

Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 8:175–185

Felix JDM, Papini-Terzi FS, Rocha FR, Vêncio RZN, Vicentini R, Nishiyama-Jr MY, Ulian EC, Souza GM, Menossi M (2009) Expression profile of signal transduction components in a sugarcane population segregating for sugar content. Trop Plant Biol 2:98–109

Gupta V, Raghuvanshi S, Gupta A, Saini N, Gaur A, Khan MS, Gupta RS, Singh J, Duttamajumder SK, Srivastava S, Suman A, Khurana JP, Kapur R, Tyagi AK (2010) The water-deficit stress- and red-rot-related genes in sugarcane. Funct Integr Genomics 10:207–214

Halford NG, Hey SJ (2009) Snf1-related protein kinases (SnRKs) act within an intricate network that links metabolic and stress signalling in plants. Biochem J 419:247–259

Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. Genome Res 9:868–877

Iseli C, Jongeneel CV, Bucher P (1999) ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. Proc Int Conf Intell Syst Mol Biol 7:138–148

Jannoo N, Grivet L, Chantret N, Garsmeur O, Glaszmann JC, Arruda P, D'Hont A (2007) Orthologous comparison in a gene-rich region among grasses reveals stability in the sugarcane polyploid genome. Plant J 50:574–585

James EK, Reis VM, Olivares FL,Baldani JI, Dobereiner J (1994) Infection of sugar cane by the nitrogen-fixing bacterium Acetobacter diazotrophicus. Journal of Experimental Botany 45: 757–766

Ma HM, Schulze S, Lee S, Yang M, Mirkov E, Irvine J, Moore P, Paterson A (2004) An EST survey of the sugarcane transcriptome. TAG Theor Appl Genet 108:851–863

Mardis ER (2008) Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet 9:387–402

Masoudi-Nejad A, Tonomura K, Kawashima S, Moriya Y, Suzuki M, Itoh M, Kanehisa M, Endo T, Goto S (2006) EGassembler: online bioinformatics service for large-scale processing, clustering and assembling ESTs and genomic DNA fragments. Nucleic Acids Res 34:W459–W462

McCormick AJ, Cramerd MD, Watt DA (2008) Regulation of photosynthesis by sugars in sugarcane leaves. J Plant Physiol 165:1817–1829

Miller JR, Koren S, Sutton G (2010) Assembly algorithms for next-generation sequencing data. Genomics 95:315–327

Nogueira FT, Rosa VE, Menossi M, Ulian EC, Arruda P (2003) RNA expression profiles and data mining of sugarcane response to low temperature. Plant Physiol 132:1811–1824

Papini-Terzi FS, Rocha FR, Vêncio RZ, Felix JM, Branco DS, Waclawovsky AJ, Bem LED, Lembke CG, Costa MD, Nishiyama-Jr MY, Vicentini R, Vincentz MG, Ulian EC, Menossi M, Souza GM (2009) Sugarcane genes associated with sucrose content. BMC Genomics 10:1–21

Papini-Terzi FS, Rocha FR, Vencio RZN, Oliveira KC, Felix JM, Vicentini R, Rocha CS, Simoes ACQ, Ulian EC, Mauro SMZ, Silva AM, Pereira CAB, Menossi M, Souza GM (2005) Transcription profiling of signal transduction-related genes in sugarcane tissues. DNA Res 12:27–38

Park SJ, Huang Y, Ayoubi P (2006) Identification of expression profiles of sorghum genes in response to greenbug phloemfeeding using cDNA subtraction and microarray analysis. Planta 223:932–947

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob-

ur-Rahman WD, Westhoff P, Mayer KFX, Messing J, Rokhsar DS (2009) The Sorghum bicolor genome and the diversification of grasses. Nature 457:551–556

Pratt LH, Liang C, Shah M, Sun F, Wang H, Reid SP, Gingle AR, Paterson AH, Wing R, Dean R, Klein R, Nguyen HT, Ma HM, Zhao X, Morishige DT, Mullet JE, Cordonnier-Pratt MM (2005) Sorghum expressed sequence tags identify signature genes for drought, pathogenesis, and skotomorphogenesis from a milestone set of 16,801 unique transcripts. Plant Physiol 139:869–884

Remm M, Storm CE, Sonnhammer EL (2001) Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. J Mol Biol 314:1041–1052

Rocha FR, Papini-Terzi FS, Nishiyama MY Jr, Vêncio RZ, Vicentini R, Duarte RD, Rosa-Jr VE, Vinagre F, Barsalobres C, Medeiros AH, Rodrigues FA, Ulian EC, Zingaretti SM, Galbiatti JA, Almeida RS, Figueira AV, Hemerly AS, Silva-Filho MC, Menossi M, Souza GM (2007) Signal transduction-related responses to phytohormones and environmental challenges in sugarcane. BMC Genomics 8:1–22

Rosenbloom KR, Dreszer TR, Pheasant M, Barber GP, Meyer LR, Pohl A, Raney BJ, Wang T, Hinrichs AS, Zweig AS, Fujita PA, Learned K, Rhead B, Smith KE, Kuhn RM, Karolchik D, Haussler D, Kent WJ (2010) ENCODE whole-genome data in the UCSC Genome Browser. Nucleic Acids Res 38:D620–D625

Salzman RA, Brady JA, Finlayson SA, Buchanan CD, Summer EJ, Sun F, Klein PE, Klein RR, Pratt LH, Cordonnier-Pratt M-M, Mullet JE (2005) Transcriptional profiling of sorghum induced by methyl jasmonate, salicylic acid, and aminocyclopropane carboxylic acid reveals cooperative regulation and novel gene responses. Plant Physiol 138:352–368

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74:5463–5467

Souza GM, Simoes ACQ, Oliveira KC, Garay HM, Fiorini LC, FdS G, Nishiyama-Junior MY, Silva AM (2001) The sugarcane signal transduction (SUCAST) catalogue: prospecting signal transduction in sugarcane. Genet Mol Biol 24:1–4

Sugden C, Donaghy PG, Halford NG, Hardie G (1999) Two SNF1-related protein kinases from spinach leaf phosphorylate and inactivate 3-hydroxy-3-methylglutaryl-coenzyme a reductase, nitrate reductase, and sucrose phosphate synthase in vitro. Plant Physiol 120:257–274

Sung-Jin P, Huang Y, Ayoubi P (2006) Identification of expression profiles of sorghum genes in response to greenbug phloem-feeding using cDNA subtraction and microarray analysis. Planta 223:932–947

Swaminathan K, Alabady M, Varala K, De Paoli E, Ho I, Rokhsar DS, Ming R, Green PJ, Meyers BC, Moose SP, Hudson ME, Arumuganathan AK (2010) Genomic and small RNA sequencing of Miscanthus x giganteus shows the utility of sorghum as a reference genome sequence for Andropogoneae grasses. Genome Biol 11:R12

Torosera D, Athwalb GS, Huber SC (1998) Site-specific regulatory interaction between spinach leaf sucrose-phosphate synthase and 14-3-3 proteins. FEBS Lett 435:110–114

Vêncio RZN, Koide T (2005) HTself: self–self based statistical test for low replication microarray studies. DNA Res 12:211–214

Vettore AL, Silva FR, Kemper EL, Arruda P (2001) The libraries that made SUCEST. Genet Mol Biol 24:1–7

Vettore AL, da Silva FR, Kemper EL, Souza GM, da Silva AM, Ferro MIT, Henrique-Silva F, Giglioti ÉA, Lemos MVF, Coutinho LL, Nobrega MP, Carrer H, França SC, Maurício Bacci J, Goldman MHS, Gomes SL, Nunes LR, Camargo LEA, Siqueira WJ, Sluys M-AV, Thiemann OH, Kuramae EE, Santelli RV, Marino CL, Targon MLPN, Ferro JA, Silveira HCS, Marini DC, Lemos EGM, Monteiro-Vitorello CB, Tambor JHM, Carraro DM, Roberto PG, Martins VG, Goldman GH, da Oliveira RC, Truffi D, Colombo CA, Rossi M, da Araujo PG, Sculaccio SA, Angella A, Lima MMA, Vicente E, de Rosa J, Siviero F, Coscrato VE, Machado MA, Grivet L, Mauro SMZD, Nobrega FG, Menck CFM, Braga MDV, Telles GP, Cara FAA, Pedrosa G, Meidanis J, Arruda P (2003) Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. Genome Res 13:2725–2735

Zhu-Salzman K, Salzman RA, Ahn J-E, Koiwa H (2004) Transcriptional regulation of sorghum defense determinants against a phloem-feeding aphid. Plant Physiol 134:420–431

Chapter 7
Sorghum and Sugarcane Proteomics

Bongani Kaiser Ndimba and Rudo Ngara

Abstract Proteomics is recognised as an important tool for global, translational level, gene expression studies. In comparison to animal studies, plant proteomics lags far behind. This chapter gives a brief review of the current state of plant proteomics. This is followed by a report of the application of in silico bioinformatics approaches in *Saccharum* spp. (sugarcane) protein subcellular localisation studies. The first comprehensive attempt at Saccharinae proteome work was done on *Sorghum bicolor*, commonly referred to as sorghum. Data from two-dimensional polyacrylamide gel electrophoresis (2D PAGE)- and mass spectrometric (MS)-based proteomics tools used to analyse global protein accumulation profiles of leaf, sheath and root tissues from two sorghum varieties, AS6 and MN1618, is described. Identified sorghum proteins are grouped into appropriate functional categories and their subcellular localisations are predicted using various bioinformatic tools.

Keywords Sorghum bicolor • Saccharinae • Proteomics • Mass spectrometry • 2D SDS-PAGE • Isoelectric focusing • MN1618 • AS6

B.K. Ndimba (✉)
Proteomics Research Group, Department of Biotechnology,
University of the Western Cape, P. Bag X17, Bellville 7535, South Africa

Proteomics Research and Services Unit, Biotechnology Platform,
Agricultural Research Council,ARC Infruitec-Nietvoorbij
Lelie Street, Idas Valley, Private Bag X5026, Stellenbosch 7599, South Africa
e-mail: bndimba@uwc.ac.za; NdimbaB@arc.agric.za

R. Ngara
Proteomics Research Group, Department of Biotechnology, University of the Western Cape,
P. Bag X17, Bellville 7535, South Africa
e-mail: rngara@uwc.ac.za

A.H. Paterson (ed.), *Genomics of the Saccharinae*, Plant Genetics and Genomics:
Crops and Models 11, DOI 10.1007/978-1-4419-5947-8_7,

1 Introduction

Proteomics is defined as the large-scale analysis of proteins from a particular organism, tissue or cell (Blackstock and Weir 1999; Pandey and Mann 2000; van Wijk 2001), while the proteome is the expressed *prote*in complement of the gen*ome* (Blackstock and Weir 1999). Unlike the genome that is generally well defined and static over time, the proteome is dynamic and constantly changes during development and in responses to both internal and external cues (Heazlewood and Millar 2003; Komatsu 2006; Speicher 2004).

The field of proteomics is divided into three broad areas: expressional, functional and structural proteomics (Blackstock and Weir 1999; Ng and Ilag 2002). Expressional proteomics (also termed differential proteomics) is the study of global changes in protein expression; functional proteomics include large-scale studies of protein–protein interactions and enzyme activity assays, while structural proteomics is the study of the three-dimensional structure of proteins. For further details on the basic principles of proteomics and its applications and technology foresight, we recommend a book entitled *Proteomics in Cancer Research* (Liebler 2004). The present chapter presents an overview of expressional proteomics and details of its application in sugarcane (*Saccharum* spp.) and sorghum (*Sorghum bicolor* L. Moench.) research.

Essential parts of the workflow in high-throughput expressional proteomics include the separation of proteins in a crude extract which often contains complex proteome and their proteolytic digestion followed by mass spectrometry-based protein identification. Crude protein extracts are separated either by one- or two-dimensional gel electrophoresis (1DE or 2DE) in gel-based proteomics or proteolysed then separated using multi dimensional liquid chromatography (MUDPIT) also known as non-gel-based (Washburn et al. 2001). Two-dimensional gel electrophoresis remains the method of choice for separating proteins in complex mixtures (Dunn and Gorg 2001; Rabilloud 2002; Thiellement et al. 1999) because of its ability to separate dozens, hundreds or even thousands of proteins at a time (Gorg et al. 2000) as well as different protein isoforms (Carpentier et al. 2008). This electrophoresis system separates proteins in two dimensions: firstly on the basis of their isoelectric point (p*I*) by isoelectric focusing (IEF); and secondly on the basis of their molecular weight (MW) using sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE). After 2DE, separated proteins are visualised in the gels using protein stains and subsequently identified by mass spectrometry (MS) and database searches.

A number of protein staining methods are available, which allow for the detection and visualisation of protein bands or spots in 1DE or 2DE gels (Gorg and Weiss 2004; Hurkman and Tanaka 2007; Westermeier 2005). The criteria used for selecting which stain to use and when depend on the ease of use, reliability, sensitivity, compatibility with MS and cost (Hurkman and Tanaka 2007). Widely used stains include Coomassie Brilliant Blue (CBB) dyes (R-250 and G-250); improved silver stains, which use sodium thiosulphate instead of aldehyde-based sensitizers that are

not compatible with MS; and fluorescent stains such as cyanine-based dyes (Cydye), SYPRO Ruby (Gorg and Weiss 2004; Hurkman and Tanaka 2007; Patton 2000; Westermeier 2005) and the Flamingo™ fluorescent stain developed by Bio-Rad Laboratories (Bio-Rad, Hercules, CA, USA).

After the detection and visualisation of proteins in 2DE gels, the gels are scanned, imaged and comparatively analysed using dedicated 2DE analysis software packages such as PDQuest, Melanie, Z3, Z4000, Phoretix and Progenesis amongst others (reviewed by Marengo et al. 2005). The analysis of 2DE gels using PDQuest software, for instance, follows a series of steps for gel evaluation, namely, scanning, image filtering, automated spot detection, matching of spot profiles, normalisation and differential and statistical analysis of the expressed proteins (Marengo et al. 2005). Thereafter, proteins of interest are proteolysed, usually by trypsin, and identified using MS-based identification methods.

Mass spectrometry, defined as the accurate mass measurement of charged analytes (Patterson and Aebersold 2003), is widely used for the identification of proteins in proteomics. Mass spectrometers have three main components: an ionization source, a mass analyser and an ion detector (Patterson et al. 2001); are named on the basis of their ionization source and mass analysers (Patterson 2000); and measure mass-to-charge ratios (m/z) of charged molecules. The two commonly used ionisation methods in proteomics are matrix-assisted laser desorption/ionisation (MALDI) and electrospray ionisation (ESI). MALDI is commonly combined with the time-of-flight (TOF) mass analyser, while ESI is combined with triple quadrupoles (Q) followed by a TOF-mass analyser (Matthiesen and Mutenda 2007). The principles and applications of these MS instruments together with a range of others are reviewed elsewhere (Aebersold and Goodlett 2001; Lin et al. 2003; Matthiesen and Mutenda 2007; Patterson 2004; Patterson and Aebersold 2003; Westermeier 2005).

2 Plant Proteomics

The field of proteomics is gaining momentum in plant sciences with several studies having been performed and reported on agriculturally important crops (Jorrin et al. 2007; Jorrin-Novo et al. 2009; Salekdeh and Komatsu 2007). Although there has been major proteomic advances using other plant species, much of the knowledge gained on plant developmental processes and stress response mechanisms has been from work using Arabidopsis (*Arabidopsis thaliana*) and rice (*Oryza sativa*) (Jorrin et al. 2007; Jorrin-Novo et al. 2009) mainly because of their publicly available complete genome sequences (International rice genome sequencing project. 2005; The Arabidopsis Genome Initiative. 2000).

In proteomics studies, genome sequences are important resource tools for the identification of proteins. Where fully annotated sequences are not yet available, protein identification can be through similarity searches of homologous proteins in closely related species (Carpentier et al. 2008). Indeed, prior to the completion of the maize genome-sequencing project (Schnable et al. 2009), a proteomic study in

maize (*Zea mays*) (Porubleva et al. 2001) relied largely on the limited sequence data and homology-based protein identification. The sugarcane proteome prediction study by Vicentini and Menossi (2009) relied on the use of expressed sequence tags (ESTs) data, which represents partial gene sequences. However, for the other plant species without significant amount of published genomic sequence or EST data, protein identification success rates are lowered resulting in limited proteomic data being available (Jorrin et al. 2007). Amongst the cereal crops, only rice (International rice genome sequencing project. 2005), sorghum (Paterson et al. 2009) and maize (Schnable et al. 2009) have been fully sequenced to date. These genome sequences offer invaluable tools for the identification of genes and proteins with potential application in plant breeding approaches for both the increase in yield as well as tolerance to both abiotic and biotic stresses (Salekdeh and Komatsu 2007).

In this chapter, we give an overview of proteomics studies in sugarcane (*Saccharum* spp.) and sorghum (*Sorghum bicolor* L. Moench.).

3 Sugarcane Proteomics

The first published work on sugarcane proteomics reported an 18 kDa protein spot seen accumulating in 2DE separated proteins that were prepared from K86-161 sugarcane plant leaves subjected to drought stress conditions (Jangpromma et al. 2007). This 18 kDa protein was not identified, but its expression under drought stress conditions was further tested and confirmed via Western blot analysis.

Recently, an in silico genome-wide proteome analysis was conducted as the first attempt in the global study of the sugarcane (*Saccharum* ssp.) proteome (Vicentini and Menossi 2009). This study used sugarcane EST sequences from a SUCEST (Sugarcane EST) project (Vettore et al. 2003) which had a collection of 43,141 sugarcane cDNA sequences, estimated to represent 30,000 genes. Vicentini and Menossi (2009) used the SUCEST data to determine open reading frames (ORFs) using three prediction programmes (GeneMark.SPL, GENSCAN and ESTScan) and these ORFs were subsequently used to construct putative protein sequences. This exercise resulted in the prediction of 11,882 putative polypeptides.

The first thing done with this putative protein data was to conduct a thorough putative subcellular localisation of the putative proteins using PWMSubLoc, a relatively novel method that combines the power of seven algorithms, namely—MitoProtII (Claros and Vincens 1996), PSORT (Nakai and Horton 1999), TargetP (Emanuelsson et al. 2000), PredictNSL (Cokol et al. 2000), SubLoc (Hua and Sun 2001), iPSORT (Bannai et al. 2002) and Predotar (Small et al. 2004). The mitochondria, plastids, secretory pathway, nucleus and cytoplasm are five subcellular compartments chosen for this investigation.

The study further analysed putative functions of proteins that were predicted to reside in sugarcane plastid and mitochondria. Polypeptides homologous to proteins with known functions were reported. These include electron transport-, metabolism-, proteolytic/peptidolytic-, protein biosynthesis-, carbohydrate metabolism-, secretory-,

photosynthetic-, protein folding-, proton transport-, DNA repair- and signal transduction-related proteins. Interestingly, bioinformatic predictions show that plastid and mitochondria have similar proportions of putative functional proteins. In both cases, electron transport-related proteins represent the majority and proton transport-related proteins are in the minority. It is also noteworthy that some of the proteins common to both compartments are secretory pathway-related proteins. This observation is important but its significance is not explained, and thus would require further study as this may be outside the scope of the currently reported work.

The impact of sugarcane on global economics is enormous as it cultivated in over 110 countries on approximately 20 million hectares and producing approximately 1,590 million metric tons of biomass per annum (http://en.wikipedia.org, and the Food and Agriculture Organisation (FAO)). Advances in the cost and speed of genome sequencing (Podolak 2010) will eventually lead to a better understanding of its genome and facilitate further studies on this plant's proteomics and related molecular biology towards further development of tools and methods for the improvement of its productivity.

4 Sorghum Proteomics

Sorghum's natural drought tolerance together with its recent genome sequencing (Paterson et al. 2009) makes it the most logical Saccharinae model for both proteomics and genomics research, especially for studies aimed at understanding the mechanisms of drought tolerance in cereals. In addition, sorghum exhibits a much greater tolerance to salt stress, in comparison to maize (Krishnamurthy et al. 2007) for instance. Sorghum is an important "fail safe" crop in the hot, dry and relatively saline regions of the world, where it provides food, feed and fuel supplies for millions of people. Despite the economic potential of this crop and the promising technique of proteomic approaches in understanding plant biological systems, to our knowledge, sorghum proteomics is still very limited.

To our knowledge, the first sorghum proteome analysis was carried out in our laboratory at the University of the Western Cape (UWC) in Cape Town, South Africa (Ndimba and Thomas 2008; Ngara et al. 2008). In that published research work (Ngara et al. 2008), we aimed at establishing a sorghum cell suspension culture system for subsequent use in the analysis of both cellular and secreted proteins during the developmental processes of this crop, as well as in response to various abiotic stresses. The use of cell suspension cultures in sorghum proteomics was largely motivated by the wide application of plant cell cultures in proteomics; the large supply of homogenous plant material provided for by these cultures; as well as the ease with which these cultures may be manipulated under a range of experimental conditions. In the study (Ngara et al. 2008), we established a viable cell culture system and profiled the 2DE protein patterns of the total soluble proteins (TSPs) and the secreted proteins, also termed culture filtrate (CF) protein. Both 1DE and 2DE separation of the TSP and CF protein extracts showed that the two

proteomes have distinct expression profiles, showing differences in nature (MW and p*I*), composition and complexity. Because of the less complex nature of the secreted protein profiles, we opted to begin further work on the secretome: both as a mapping exercise as well as in studying its responsiveness towards abiotic stresses. This component of the sorghum secretome research work is still ongoing in our research group at UWC and we hope it will provide a foundation from which future studies on sorghum secretomics may spin off.

Our research group is also involved in the application of sorghum whole plant systems in proteomics studies. This approach allows for the integration of tissue-specific proteome profiling and hence in-depth analysis of biological processes simultaneously at work in different locations of the plant. Outlined below is a review on the optimisation of the extraction of TSPs from various tissues of two sorghum varieties (AS6 and MN1618). These varieties were part of a seed batch that was donated to us by Dr. Pangirayi Tongoona, a plant breeder at the African Centre for Crop Improvement, University of KwaZulu-Natal, South Africa, and were chosen on the basis of their differences in salt tolerance as observed in an experiment on screening of sorghum varieties for salt tolerance (data not shown). The overall aim of this work was to establish the proteome profiles of the tissues by 2DE and work towards the establishment of a sorghum leaf proteome reference map by a combination of 2DE and mass spectrometry. The work reviewed below was part of Rudo Ngara's PhD thesis obtained at UWC (Ngara 2009).

4.1 Protein Extraction from Different Tissues of Sorghum Seedlings

Two sorghum varieties, AS6 and MN1618, were used as sources of plant material. The sorghum seeds were surface sterilised and germinated on Murashige and Skoog (MS) media [4.4 g/l MS basal medium (Murashige and Skoog 1962), 3% (w/v) sucrose, 5 mM 2-(*N*-morpholino)ethanesulfonic acid (MES) and 0.8% (w/v) agar adjusted to pH 5.8 using 2 M KOH]. Fourteen days following seed plating, sorghum seedlings were harvested and the leaf, sheath and root tissues were excised. To minimise protein degradation, the plant material was immediately flash frozen in liquid nitrogen and stored at −20°C until needed for protein extraction procedures. The different tissues were separately extracted of composite proteins using the TCA/acetone method as described (Ngara 2009). Leaf and sheath protein extracts were prepared from an average of ten 14-day-old sorghum seedlings. To bulk up root material for protein extraction, root protein extracts were prepared from at least 20 14-day-old sorghum seedlings. Extracted proteins were quantified using a modified Bradford assay (Bradford 1976) as previously described (Ndimba et al. 2003).

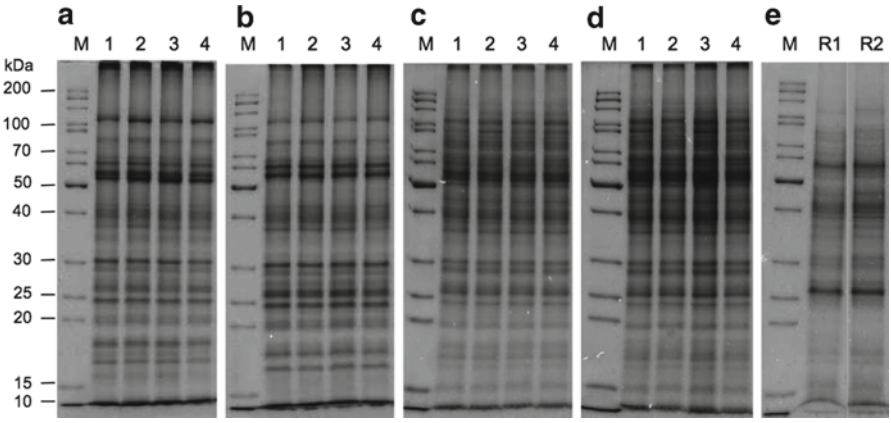

Fig. 7.1 One-dimensional gel electrophoresis of total soluble protein of leaf, sheath and root tissues of AS6 and MN1618 sorghum varieties. Approximately 10 μg of soluble protein of each tissue was loaded onto 12% SDS-PAGE gels. (**a**) AS6 leaf, (**b**) MN1618 leaf, (**c**) AS6 sheath, (**d**) MN1618 sheath and (**e**) AS6 and MN1618 roots samples. Lane *M* is the molecular weight marker. Lanes *1–4* represent independent biological replicate protein extracts for the leaf and sheath samples (**a**–**d**). Lanes *R1* and *R2* represent root samples from AS6 and MN1618 sorghum varieties, respectively

4.2 One-Dimensional Protein Profiles of Sorghum Tissues

1DE was carried out to evaluate the quality and loading quantities of the protein extracts prior to 2DE. Figure 7.1 shows CBB-stained 1DE profiles of leaf, sheath and root proteomes of the two sorghum varieties. Figure 7.1a: AS6 leaf; b: MN1618 leaf; c: AS6 sheath; d: MN1618 sheath; e: AS6 and MN1618 root proteomes. Lane M shows the MW markers. Lanes 1–4 show protein profiles from four independent biological replicate extractions for the leaf and sheath tissues. Each lane was loaded with approximately 10 μg of total protein of the respective tissue extract. It was observed that the quality of all protein extracts was good, showing no visible signs of streaking and protein degradations. The biological replicates (Lanes 1–4) within an experiment (Fig. 7.1) also showed high similarity in protein expression, abundance and banding patterns. This suggests that protein preparation was reproducible between independent extractions. Figure 7.1e illustrates the 1DE protein profiles of root tissue. Lanes R1 and R2 represent protein extracts from AS6 and MN1618 sorghum varieties, respectively. Both lanes were loaded with 10 μg of total protein. When compared to the leaf and sheath proteomes (Fig. 7.1a–d), the root proteome profiles of both seed types (Fig. 7.1e) were composed of mostly low-abundant protein bands. However, protein extracts from all three tissues, leaf, sheath and root, covered the MW range between 10 and 110 kDa.

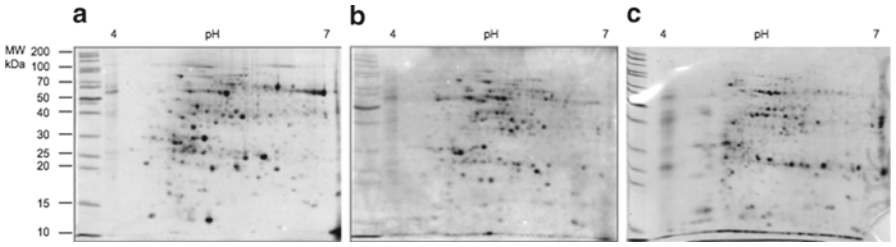

Fig. 7.2 Representative CBB-stained 2DE gels of leaf, sheath and root tissues of AS6 sorghum variety. Protein extracts were separated in the first dimension by isoelectric focusing using 7 cm linear IPG strips, pH range 4–7 and 12% SDS-PAGE gels in the second dimension. Protein loading was (**a**) leaf (100 μg), (**b**) sheath (150 μg) and (**c**) root (100 μg)

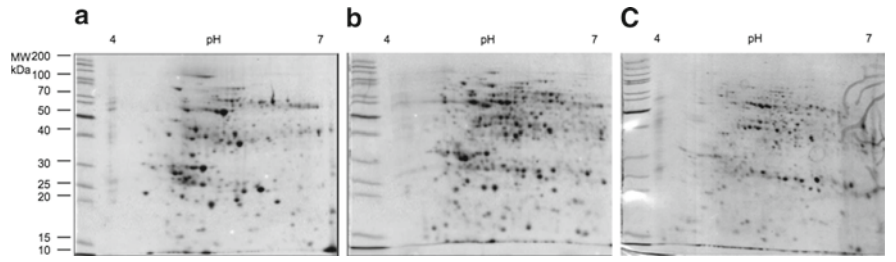

Fig. 7.3 Representative CBB-stained 2DE gels of leaf, sheath and root tissues of MN1618 sorghum variety. Protein extracts were separated in the first dimension by isoelectric focusing using 7 cm linear IPG strips, pH range 4–7 and 12% SDS-PAGE gels in the second dimension. Protein loading was (**a**) leaf (100 μg), (**b**) sheath (150 μg) and (**c**) root (100 μg)

4.3 Two-Dimensional Protein Profiles of Sorghum Tissues

Three of the four biological replicates of protein extracts from leaf and sheath tissues (Lanes 1–4, Fig. 7.1a–d) were randomly selected for further separation using 2DE analysis. Due to the limitations in protein quantities for both AS6 and MN1618 root extracts, only one composite proteome gel was prepared for each of the two sorghum varieties. Protein loading on the 2DE gels also varied between tissues depending on the overall complexity and abundance of the respective proteomes. For sheath samples, 150 μg of total protein was loaded onto immobilized pH gradient (IPG) strips (Bio-Rad), while for both the leaf and root samples, a protein load of 100 μg was used. For all tissues, mini gels, using 7 cm IPG strips of pH range 4–7, were used. This pH range was chosen on the basis that most of the soluble proteins fell within that pH range (data not shown). It was observed that spot resolution and abundance between three biological replicate gels of each of the leaf and sheath extracts were uniform. This indicates that 2DE was reproducible between different samples within an experiment. Figures 7.2 and 7.3 illustrate representative 2DE gels of tissue proteomes of AS6 and MN1618 sorghum varieties, respectively. Figure 7.2a–c shows

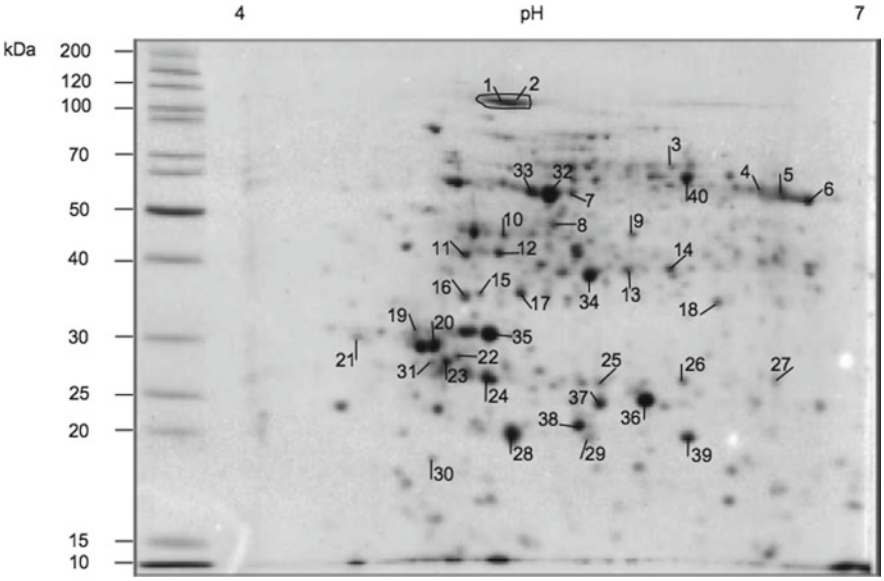

Fig. 7.4 A representative CBB-stained 2DE gel of the sorghum MN1618 leaf proteome showing spots picked for mass spectrometry analysis. About 100 μg of the leaf protein extract was separated in the first dimension by isoelectric focusing using 7 cm linear IPG strips, pH range 4–7 and 12% SDS-PAGE gels in the second dimension. Numbered spots (1–40) were selected for identification using a combination of MALDI-TOF MS, MALDI-TOF–TOF MS and database searches

the leaf, sheath and root proteomes of AS6 sorghum variety. Figure 7.3a–c represents the leaf, sheath and root proteomes of MN1618 sorghum variety. In general, it was observed that the sheath shared common spots with both the leaf and root. However, protein expression between the leaf and root was remarkably different.

4.4 Towards the Establishment of a Sorghum Leaf Proteome Map

One of the main objectives of our sorghum whole plant proteomics research was to work towards the establishment of a sorghum leaf proteome map. To achieve this goal, MN1618 leaf proteome was used in the mapping exercise. Of the 171 reproducible CBB stainable leaf protein spots observed amongst three biological replicate extractions of MN1618, a total of 40 well-resolved spots of varying degrees of abundance, MW and p*I* ranges were selected for identification using MS-based methods. The selected protein spots (spots 1–40; Fig. 7.4) were picked from CBB-stained gels using the ExQuest™ spot cutter (Bio-Rad). Protein contained in the picked gel plugs was trypsinised and peptide digests were analysed using matrix-assisted

laser desorption/ionisation-time of flight–time of flight (MALDI-TOF–TOF) MS on a 4800 Proteomic analyser (Applied Biosystems, Boston, MA, USA). Combined lists of the MS and MS–MS data were used for database searching with MASCOT version 2.2 (http://matrixscience.com) against all entries in the National Center for Biotechnology Information (NCBI) database. Although the sorghum genome has been fully sequenced (Paterson et al. 2009), and is available on the Phytozome version 5.0 database (http://www.phytozome.net/), there is still limited sorghum sequence data on databases that can be searched via the mass spectrometry data search engine MASCOT. Therefore, using sequence data on NCBI, protein identification was mainly based on the homology between sorghum and other green plants. Only spot identities with a MOWSE score equal to or greater than 76 were regarded as significant protein matches (significance threshold of $p < 0.05$).

Mass spectrometry using a combination of MALDI-TOF and MALDI-TOF–TOF MS, and genomic database searches of the 40 trypsinised protein spots resulted in the positive identification of 28 protein spots (70% successful identification of the total number of picked spots for MS identification). The remaining 12 spots did not match any protein identity with acceptable MOWSE score values. Table 7.1 gives a summary of the identified protein spots, their best match identities, plant species of origin and corresponding NCBI accession numbers. Both the theoretical and experimental MW/p*I*s of the proteins are also shown. The positively identified protein spots matched the identities of proteins from a wide range of plants possibly showing the high degree of conserved genes and gene products within higher plants. On the other hand, unidentified proteins (spots 3, 4, 15, 16, 19, 22, 23, 25, 27, 30, 31 and 38; Fig. 7.4; Table 7.1) were of varying degrees of abundance, MW and p*I* values.

4.5 Proteins Identified in Multiple Spots

It was observed that seven classes of proteins were represented in multiple spots on the 2DE gels (Fig. 7.4; Table 7.1). These are pyruvate phosphate dikinase (spots 1 and 2); RuBisCo (spots 5 and 6); malate dehydrogenase (MDH) (spots 10 and 14); ferrodoxin-NADP oxidoreductase (spots 17 and 18); adenosine diphosphate glucose pyrophosphate (spots 29 and 39); chloroplastic ATP synthase proteins (spots 7, 32, 33, 13, 40) and hydroxynitrile lyase (spots 20 and 28). The multiple protein-spotting patterns observed in this sorghum leaf proteome can be loosely classified into four groups. Group one consists of protein spots with the same NCBI accession number, migrating at the same MW but different p*I*s. Examples of proteins in this group include (1) pyruvate phosphate dikinase (spots 1 and 2; NCBI accession AAP23874; Table 7.1), which migrated at the same MW of approximately 110 kDa but different p*I*s of 5.2 and 5.3, respectively; and (2) adenosine diphosphate glucose pyrophosphate (spots 29 and 39; NCBI Accession CAC85479; Table 7.1), both migrated at 20 kDa but different p*I*s of 5.6 and 6.1, respectively. These protein spots most likely represent post-translationally modified peptides and thus isoforms of each enzymes.

Table 7.1 List of sorghum leaf proteins identified by a combination of MALDI-TOF MS, MALDI-TOF–TOF MS and database searching

Spot[a]	Best protein match	Plant species	Accession no.[b]	MOWSE score[c]	Theoretical MW/pI[d]	Experimental MW/pI[e]	Location[f]
Carbohydrate metabolism							
1	Pyruvate phosphate dikinase	*Sorghum bicolor*	AAP23874	291	102/5.7	110/5 .2	Chloroplast
2	Pyruvate phosphate dikinase	*Sorghum bicolor*	AAP23874	201	102/5.7	110/5.3	Chloroplast
5	RuBisCO, large subunit	*Stachys byzantina*	AAM33283	120	50/6.9	53/6.5	Chloroplast
6	RuBisCO, large subunit	*Anemopsis californica*	AAF14707	238	51/6.6	51/6.6	Chloroplast
9	Cytosolic 3-phosphoglycerate kinase	*Zea mays*	AAO32644	250	32/4.7	45/5.8	Cytoplasm
10	Chloroplastic NADP-dependent malate dehydrogenase	*Sorghum bicolor*	7MDH_D	121	41/4.9	45/5.2	Chloroplast
14	Cytoplasmic NAD-dependent malate dehydrogenase	*Zea mays*	AAB64290	151	36/6.0	39/6.0	Cytoplasm
11	Sedoheptulose-1,7-biphosphate precursor	*Oryza sativa*	AA022559	143	42/5.9	42/5.0	Chloroplast
12	Phosphoribulokinase, chloroplast precursor	*Mesembryanthemum crystallium*	P27774	159	44/6.4	42/5.2	Chloroplast
17	Putative ferrodoxin-NADP oxidoreductase	*Oryza sativa*	NP_910234	261	40/8.7	36/5.3	Chloroplast
18	Ferrodoxin-NADP oxidoreductase	*Zea mays*	1GAW_ B	108	39/7.3	35/6.2	Chloroplast
29	Adenosine diphosphate glucose pyrophosphatase	*Triticum aestivum*	CAC85479	137	22/6.1	20/5.6	Apoplast
39	Adenosine diphosphate glucose pyrophosphatase	*Triticum aestivum*	CAC85479	143	22/6.1	20/6.1	Apoplast
34	Fructose 1,6-biphosphate aldolase precursor	*Avena sativa*	AAF74202	269	42/9.2	39/5.6	Chloroplast
35	Oxygen-evolving enhancer protein 1 precursor	*Bruguiera gymnorrhia*	BAA96365	268	35/6.7	30/5.1	Chloroplast
36	Photosystem II oxygen-evolving complex protein 2 (fragment)	*Arabidopsis thaliana*	PA0013	112	1.4/10.2	39/5.9	Chloroplast
37	Chlorophyll *a*/*b*-binding protein type III precursor	*Lycopersicon esculentum*	1609235A	114	29/9.1	38/5.6	Chloroplast

(continued)

Table 7.1 (continued)

Spot[a]	Best protein match	Plant species	Accession no.[b]	MOWSE score[c]	Theoretical MW/p*I*[d]	Experimental MW/p*I*[e]	Location[f]
Proton transport							
7	ATP synthase subunit beta, chloroplastic	*Sorghum bicolor*	1711264A	314	54/5.1	55/5.5	Chloroplast
32	ATP synthase subunit beta, chloroplastic	*Sorghum bicolor*	1711264A	650	54/5.1	55/5.4	Chloroplast
33	ATP synthase subunit beta, chloroplastic	*Sorghum bicolor*	1711264A	213	54/5.1	55/5.3	Chloroplast
13	ATP synthase gamma chain 1, chloroplast	*Oryza sativa*	XP_478377	118	40/8.5	39/5.8	Chloroplast
40	ATP synthase CF1 alpha chain	*Oryza sativa*	AAP54723	455	56/6.1	57/6.0	Chloroplast
Protein synthesis							
8	Translational elongation factor Tu	*Oryza sativa*	XP_466527	277	50/6.6	47/5.4	Chloroplast
21	Nucleic acid-binding protein—maize	*Zea mays*		82	33/nd	30/4.5	nd
Hydrolytic enzyme							
20	Hydroxynitrile lyase	*Sorghum bicolor*	1GXS_C	260	30/4.5	30/4.9	Cytoplasm
28	Hydroxynitrile lyase	*Sorghum bicolor*	1GXS_C	285	18/4.5	20/5.1	Cytoplasm
Nucleic acid metabolism							
24	Putative adenylate kinase, chloroplast	*Oryza sativa*	XP_479721	126	32/8.1	27/5.1	Chloroplast
Detoxifying enzymes							
26	Glutathione transferase (EC 2.5.1.18) I	*Zea mays*	P12653	108	24/7.5	26/6.0	Cytoplasm
Spots with no significant matches							
	3; 4; 15; 16; 19; 22; 23; 25; 27; 30; 31; 38.						

nd not determined

[a]Spot number as indicated on the 2DE gel images (Fig. 7.4)

[b]Accession number in the National Center for Biotechnology Information (NCBI) database

[c]Combined MOWSE score for MALDI-TOF and MALDI-TOF–TOF MS

[d]Theoretical MW and p*I* were calculated on the mature peptide using the programme on http://isoelectric.ovh.org

[e]Experimental MW and p*I* were estimated from the 2DE gels shown in Fig. 7.4

[f]Subcellular location of the proteins as predicted by TargetP version 1.1 (http://www.cbs.dtu.dk/services/TargetP/); Predotar version 1.03 (http://urgi.versailles.inra.fr/predotar/predotar.html) and literature sources

Group two consists of proteins with the same NCBI Accession number, but different MW as well as p*I*s. The hydroxynitrile lyase protein (spots 20 and 28; NCBI Accession 1GXS_C; Table 7.1) could be classified in this group. These spots migrated at 30 and 20 kDa and had p*I* values of 4.9 and 5.1, respectively. According to Wajant and Mundry, *S. bicolor* hydroxynitrile lyase is composed of two different subunits α and β with molecular sizes of 33 and 22 kDa, respectively (Wajant and Mundry 1993). These subunits occur as αβ dimers (MW = 55 kDa) while the active enzyme is a heterotetramer ($\alpha_2\beta_2$; MW range 95–105 kDa). Therefore, the two hydroxynitrile lyase spots observed in this sorghum leaf proteome could probably represent α and β subunits of the enzyme. Hydroxynitrile lyases are also documented to have p*I* ranges of between 4 and 5.5 (Poulton 1990), a characteristic consistent with the results obtained in this leaf proteome study (Table 7.1).

Group three consists of spots that match the same protein name but of different NCBI Accessions, MW and p*I* values. Examples of these proteins include MDH (spots 10 and 14; Table 7.1) and ferrodoxin-NADP oxidoreductase (spots 17 and 18; Table 7.1). These peptides are most likely to represent isoforms from a multigene family. For example, the MDH protein spots represent isoforms of different subcellular locations and co-enzyme specificity and therefore are likely to be involved in different metabolic functions in the cell. Spot 10 represents a chloroplastic NADP-dependent MDH while spot 14 represents that of a cytoplasmic specific NAD-dependent MDH (Table 7.1).

Group four consists of spots with a combination of parameters from the other three groups and are also known to be part of protein–protein complexes. Examples are (1) the chloroplastic ATP synthase proteins (spots 7, 13, 32, 33 and 40; Table 7.1). These spots corresponded to the alpha (spot 40), beta (spots 7, 32 and 33) and gamma (spot 13) subunits of the chloroplastic ATP synthase complex; and (2) RuBisCo large subunits (spots 5 and 6; Table 7.1). The identification of component subunits of protein complexes has also been reported in other proteomic studies. Albertin and co-workers identified six spots corresponding to the alpha, beta, delta and gamma subunits of the chloroplastic ATP synthase complex in *B. napus* stem tissue (Albertin et al. 2009). In the same study, both the small and large subunits of RuBisCo were identified in stem tissue, while alpha and beta subunits of tubulin were detected in both the stem and root tissues.

5 Putative Functional Classification of Identified Sorghum Leaf Proteins

After the identification of expressed proteins in a particular tissue, it is important to establish their functions. Knowledge on protein function would lead to the identification of cellular processes, main metabolic pathways and biological functions of the tissue under study. The putative functions of the identified sorghum leaf proteins were assessed by a combination of similarity searches on the Arabidopsis database (http://www.arabidopsis.org), Universal Protein Sequence database (http://www.uniprot.org)

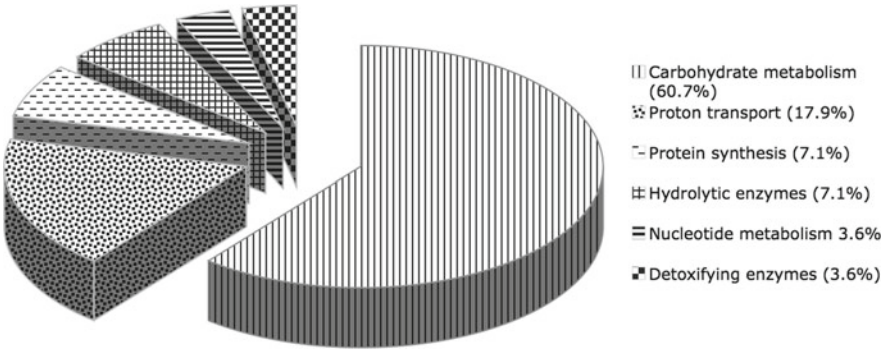

Fig. 7.5 Functional distribution of the mass spectrometry identified sorghum leaf proteins. *Numbers* indicated in *brackets* represent the proportion of proteins within each functional category expressed as a percentage of the 28 MALDI-TOF and MALDI-TOF–TOF MS positively identified protein spots

and literature sources. Using these bioinformatics tools and literature sources, all the 28 positively identified protein spots (Table 7.1) were successfully classified into six broad functional categories: carbohydrate metabolism, proton transport, protein synthesis, hydrolytic enzymes, nucleotide metabolism and detoxifying enzymes. The functional categories and proteins in each respective class are listed in Table 7.1 while a graphical representation of the distribution of proteins in each class is illustrated in Fig. 7.5. A description of proteins and their functions in each of the functional categories is given below.

5.1 Carbohydrate Metabolism

The majority of the identified sorghum leaf proteins (60.7%, Fig. 7.5) had functions in carbohydrate metabolism. This observation is consistent with results obtained in other leaf proteomics studies of maize (Porubleva et al. 2001), barrel medic (Watson et al. 2003), rice (Nozu et al. 2006) and pea (Schiltz et al. 2004). In this functional category, many biological processes and metabolic pathways including the light reaction of photosynthesis, the Calvin cycle, starch biosynthesis, glycolysis and malate/oxaloacetate shuttle system among others were represented. The high representation of such metabolic pathways in the leaf tissue of seedlings demonstrates their important contribution in the growth and development of plants as well as the primary function of plant leaves.

Five proteins that are directly involved in light reactions of photosynthesis were identified (Fig. 7.4; Table 7.1). These include the chlorophyll *a*/*b* binding proteins type II precursor (spot 37), the oxygen evolving enhancer protein 1 precursor (spot 35), the photosystem II oxygen evolving complex protein 2 (spot 36) and ferrodoxin NADP-oxidoreductase (spots 17 and 18). Chlorophyll-binding proteins are synthesised

as precursor molecules in the cytoplasm and imported into the chloroplast where they are inserted in the thylakoid membranes (Bassi et al. 1997). They have several functions including light harvesting, energy dissipation and pigment storage. As components of the light harvesting complexes in plants, the primary functions of chlorophyll *a*/*b* binding proteins is the absorption of light and the transfer of the excitation energy to the photochemical reaction centres (Bassi et al. 1997; Ganeteg et al. 2001). In some cases, plants are exposed to higher light intensities than used in photosynthesis. Therefore, to prevent photoinhibition and damage to the photosynthetic machinery excess energy is then dissipated by these light-harvesting proteins. In addition, chlorophyll *a*/*b* binding proteins are believed to have a function in pigment storage (Bassi et al. 1997).

The light energy absorbed by chlorophyll *a*/*b* binding proteins (spot 37; Table 7.1) is used to drive the light-dependent oxidation of water, releasing molecular oxygen. Hydrogen ions are also released in the process, creating a transmembrane chemiosmotic potential that is utilised by adenosine triosephosphate (ATP) synthases during ATP synthesis. Photolysis of water occurs in the oxygen-evolving complex (OEC) of Photosystem II (PS II) reaction centres (McEvoy and Brudvig 2006; Raymond and Blankenship 2008; Sproviero et al. 2007). The OEC is composed of four manganese ions, calcium and possibly chloride ions, which are bound to extrinsic proteins (McEvoy and Brudvig 2006). Two of these extrinsic OEC proteins were identified in our sorghum leaf proteome study: the PS II OEC protein 2 (spot 36; Table 7.1) and an oxygen-evolving enhancer protein precursor (spot 35; Table 7.1). PS II OEC proteins are involved in retaining calcium and chloride ions, two inorganic cofactors for the water-splitting reaction (Ifuku et al. 2005). The oxygen-evolving enhancer protein is believed to have a dual function: (1) optimising the manganese cluster during photolysis and (2) protecting the reaction centre proteins from damage by oxygen radical formed in light (Heide et al. 2004). Two isoforms of ferrodoxin-NADP oxidoreductases (spots 17 and 18; Table 7.1), with MW of approximately 36 and 35 kDa, respectively, were identified in this sorghum leaf proteome. The MW range of these spots is consistent with the average 35 kDa of plant reductases (Arakaki et al. 1997). In plants, this enzyme exists in two different forms: photosynthetic and heterotrophic forms, which are encoded for by different genes and may be associated with different metabolic pathways (Gummadova et al. 2007). In our sorghum leaf proteomics study, only the photosynthetic isoforms were identified. Ferrodoxin-NADP oxidoreductases catalyse the reversible electron transfer between one-electron carrier systems (ferrodoxin) and the two-electron carrying NADP(H) (Thomas et al. 2006). In chloroplasts, the main physiological function of this enzyme is to catalyse the final step of the photosynthetic electron transport, providing NADPH, which is then utilised in the carbon fixation step of the Calvin cycle (Arakaki et al. 1997).

The Calvin cycle (also termed the reductive pentose phosphate pathway) is a metabolic pathway that produced pentose sugars (Heldt 1997). The cycle is characterised by three phases: the carboxylation, reduction and regeneration phases. Some proteins involved in these phases were identified in our sorghum leaf proteome study. Two spots (5 and 6; Table 7.1) representing the RuBisCO large subunits were

identified. RuBisCO is a multimeric enzyme with two subunits: large (50–55 kDa) and small (12–18 kDa) (Andersson and Backlund 2008). In this study, the RuBisCO proteins were observed as forming a train of spots on the basic side of the gel (p*I*s 6.5 and 6.6) with MW of approximately 53 and 51 kDa, respectively. This observation is consistent with results from other proteomic studies. In the pea leaf proteome, it was observed that RuBisCO proteins formed an abundant train of spots between pH 6 and 7, at approximately 50 kDa (Schiltz et al. 2004). Similarly, in the maize leaf proteome, several RuBisCO large subunits were also identified between pH 6 and 7 and MW of approximately 50–56 kDa (Porubleva et al. 2001). However, contrary to reports that RuBisCO may contribute to the low quality of 2DE gels and also obscure the resolution of other relatively low-abundant proteins (Watson et al. 2003), in our study, the expression of the large subunit of RuBisCo did not seem to compromise the quality of our sorghum leaf 2DE gel images using 7 cm, IPG strips of pH range 4–7 (Bio-Rad).

Functionally, RuBisCO catalyses the carbon fixation (carboxylation) reaction in the Calvin cycle of photosynthesising plants. In this process, ribulose 1,5-bisphosphate (RuBP), a 5-carbon compound, serves as an acceptor molecule for CO_2 to form an unstable 6-carbon compound, which immediately breaks down, forming two molecules of 3-phosphoglycerate (3PGA) (Andersson and Backlund 2008; Kellogg and Juliano 1997; Tabita et al. 2007). The end product of this carboxylation reaction, 3PGA, is phosphorylated by ATP to form 1,3-biphosphoglycerate and adenosine diphosphate (ADP). This reaction is catalysed by a cytosolic 3-phosphoglycerate kinase also identified in this study as protein spot 9 (Table 7.1). The above reaction is one of the two that occurs in the second phase (reduction phase) of the Calvin cycle (Heldt 1997; Macdonald and Buchanan 1997).

In the third phase, RuBP molecules are regenerated to allow the first carbon fixation step to occur. The regeneration phase is characterised by a series of enzymatic reactions that convert triose phosphate to RuBP (Heldt 1997; Macdonald and Buchanan 1997). Some of the enzymes involved in the intermediate steps of this phase, a sedoheptulose-1,7-biphosphate precursor (spot 11; Table 7.1) and fructose 1,6-biphosphate aldolase precursor (spot 34; Table 7.1), were also identified. Together with others, these two enzymes catalyse reactions, which ultimately result in the formation of ribulose-5-phosphate. The ribulose-5-phosphate is then phosphorylated to form RuBP by phosphoribulokinase, an enzyme that was also identified in the sorghum leaf proteome (spot 12; Table 7.1). To complete the cycle, RuBP is subsequently used as a substrate for RuBisCo in the first phase of carbon fixation. Some of the triose phosphate produced in the Calvin cycle is used for sucrose and starch biosynthesis (Raines 2003; Tamoi et al. 2005).

Starch is an important storage polysaccharide in plants, providing an energy source for various metabolic processes (Kruger 1997). Starch synthesis involves three enzymes: adenosine diphosphate glucose pyrophosphatase (AGPase), a starch synthase and a branching enzyme (Guan and Keeling 1998; Martin and Smith 1995; Preiss 1997). Only one of these three enzymes was identified in the sorghum leaf proteome. Two protein spots (29 and 39; Table 7.1) representing an AGPase were identified, both migrating at the same MW of approximately 20 kDa but having

different p*I*s of 5.6 and 6.1, respectively. Plant AGPases are tetrameric in structure, being composed of two different subunits, which are products of different genes. The small and large subunits have a subunit MW range of 50–54 kDa and 51–60 kDa, respectively (Preiss 1997). However, although the experimental and theoretical MW of the two AGPases spots are in close range, 20 kDa versus 22 kDa, respectively, for both spots (spots 29 and 39; Table 7.1), these values were almost half of those published for general plant AGPases. This observation could possibly indicate (1) the high similarity in protein sequences between the two grasses (wheat and sorghum) and (2) differences in amino acid sequences and MW between AGPases of grasses and other plant AGPases.

AGPases catalyse the formation of ADP-glucose and inorganic pyrophosphate from ATP and glucose-1-phosphate (Boehlein et al. 2005). The end product of this reaction, ADP-glucose, is a precursor for starch synthesis (Tetlow et al. 2003). Starch synthase then transfers the glucose from ADP-glucose to the nonreducing end of a growing acceptor chain, thus elongating the α-1,4 glucan chains. In the third step, the starch branching enzyme then cleaves an elongated α-1,4 glucan chain, simultaneously transferring it to an acceptor chain to form α-1,6 linkages (Guan and Keeling 1998; Martin and Smith 1995; Preiss 1997).

Two isoforms of MDH were identified in the sorghum leaf proteome. Spot 10 represents a chloroplast NADP-dependent MDH (EC 1.1.1.82) with an approximate experimental MW of 45 kDa and p*I* of 5.2 (Table 7.1). The second isoform in spot 14 represents a cytoplasmic NAD-dependent MDH (EC 1.1.1.37) with an approximate experimental MW of 39 kDa and p*I* of 6 (Fig. 7.4; Table 7.1). Plant cells are known to contain multiple isoforms of MDHs, which differ in co-enzyme specificity, subcellular localisation and biological function (Ding and Ma 2004; Minarik et al. 2002). In plants, five different classes of MDHs are present: (1) chloroplast NADP-dependent MDH; (2) mitochondrial NAD-dependent MDH; (3) microbody NAD-dependent MDH; (4) chloroplast NAD-dependent MDH and (5) cytosolic NAD-dependent MDH (Ding and Ma 2004). These enzymes occur as homodimers, with subunit molecular weight ranging between 32 and 37 kDa for the NAD-dependent MDH and 42–43 kDa for the NADP-dependent MDHs (Ding and Ma 2004). The identification of these two isoforms of MDH with different subcellular localisation and co-enzyme specificity in sorghum leaf tissue reinforces this notion.

The experimental/theoretical MW of 45/41 kDa and 39/36 kDa for spots 10 and 14, respectively, (Table 7.1) are in close range. This possibly indicates the high degree of similarities between the amino acid sequences of the proteins identified in our sorghum leaf proteome and those found in plant databases. Furthermore, the experimental MW of spot 10, a chloroplastic NADP-dependent MDH (MW = 45 kDa; Table 7.1), is in close agreement with the average MW range (42–43 kDa) for chloroplastic NADP-dependent MDH. Similarly, the experimental MW of spot 14, a cytoplasmic NAD-dependent MDH (MW = 39 kDa; Table 7.1), is in close agreement with the average MW range (32–37 kDa) for cytoplasmic NAD-dependent MDH. With respect to the p*I*s of the two proteins, experimental and theoretical values also compare well for both spots. For example, the p*I* of spot 14 (p*I* = 5.2; Table 7.1), a cytoplasmic NAD-dependent MDH, is comparable with that of a

stromal NAD-dependent MDH isoform (p*I* = 5.3) isolated from chloroplast of spinach leaves (Cvetic et al. 2008).

Generally, MDHs catalyse the interconversion of oxaloacetate and malate using the NAD/NADP coenzyme system (Goward and Nicholls 1994; Minarik et al. 2002). However different isoforms in different subcellular locations are thought to have different functions. For example, the chloroplastic NADP-dependent MDH forms part of a malate valve system (Scheibe 2004), which converts excess NADPH into malate and transports in from the chloroplast into the cytosol (Fridlyand et al. 1998). Therefore, it is highly probable that the chloroplastic NADP-dependent MDH identified in our sorghum leaf proteome might have a function in balancing reducing equivalents between the cytosol and the chloroplast stroma. In C_4 plants such as sorghum and maize, this chloroplastic NADP-dependent MDH isoform may have an additional role in the synthesis of malate, which is transported into the chloroplast of bundle sheath cells and takes part in carbon fixation (Ding and Ma 2004). On the other hand, cytoplasmic NAD-dependent MDH isoforms are less well characterised with limited structural and functional information being known. Nevertheless, a cytoplasmic NAD-dependent MDH was isolated from wheat (TaMDH) (Ding and Ma 2004). Both mRNA and protein expression studies showed that this MDH was expressed in leaves, stems and roots. TaMDH had an MW of approximately 40 kDa, which corresponds very well with 39 kDa for protein spot 14 (Table 7.1), observed in the sorghum leaf proteome. Since the cytoplasmic NAD-dependent MDH isoforms were shown to be present in different plants tissues (Ding and Ma 2004), they are proposed to have housekeeping functions in plant metabolism. However, their actual physiological functions and mechanism of action are yet to be elucidated.

5.2 Proton Transport

The second major functional category (17.9%) identified in the sorghum leaf proteome study consisted of proteins associated with the transportation of protons across the chloroplastic thylakoid membranes. A total of five spots (spots 7, 13, 32, 33 and 40; Table 7.1) representing various subunits of the chloroplastic ATP synthase complex were identified. The relative abundances of the subunits also varied as observed in Fig. 7.4. Various subunit components of this complex have also been identified in the proteomes of rapeseed stem (Albertin et al. 2009) and maize leaf (Porubleva et al. 2001). Structurally, chloroplastic ATP synthases have two major components: an extrinsic CF_1, which synthesises ATP, and the membrane-bound CF_0 that translocates protons across the thylakoid membrane. The CF_1 has five subunits alpha, beta, gamma, delta and epsilon while CF_0 has three main subunits: a, b and c (McCarty 1992). The five protein spots identified in this sorghum leaf proteome represent three of CF_1 subunits: alpha, beta and gamma with experimental MW of 57, 55 and 39 kDa, respectively (Table 7.1). These estimated MW ranges are highly comparable to the theoretical estimates of 56, 54, and 40 kDa (Table 7.1) and

55, 54 and 36 kDa (McCarty 1992) for alpha, beta and gamma subunits, respectively. The main physiological function of the chloroplastic ATP synthase is to produce ATP from ADP in the presence of a proton gradient across the thylakoid membrane (McCarty 1992; von Ballmoos and Dimroth 2007). This proton gradient is created during the light reactions of photosynthesis. The ATP then drives a wide variety of energy-consuming cellular processes such as the Calvin cycle during plant cell growth and development.

5.3 Protein Synthesis

Two proteins, a translational elongation factor Tu (spot 8; Table 7.1) and a nucleic acid-binding protein (spot 21; Table 7.1), represented the protein synthesis group, thus constituting 7.1% of all the identified proteins. Protein synthesis is important in providing cells with the needed proteins and enzymes, which participate in many biological processes within the cell. During protein synthesis, elongation factor Tu binds aminoacyl-tRNAs and guanosine tri-phosphate (GTP) to form a complex, which then associates with a ribosome that is complexed to messenger RNA and transfer RNA (Harris et al. 1994; Kang et al. 1998). In this way, polypeptides are elongated during synthesis.

5.4 Nucleotide Metabolism

The nucleotide metabolism functional category was represented by one protein, a putative adenylate kinase (ADK) (spot 24; Table 7.1), thus constituting about 3.6% of all the identified proteins. ADKs are small monomeric enzymes of approximately 21–27 kDa in size (Schiltz et al. 1994). The enzymes catalyse the reversible formation of ADP by the transfer of one phosphate group from ATP to adenosine monophosphate (AMP) (Lange et al. 2008). As such, ADK is considered to be an important enzyme in energy metabolism as well as in maintaining the equilibrium of adenylates (ADP, ATP and AMP) in vivo (Carrari et al. 2005; Igamberdiev and Kleczkowski 2006). In maize, a C_4 plant, ADK is important for efficient CO_2 fixation in the C_4 cycle by removing and recycling AMP produced in the pyruvate phosphate dikinase reaction (Schiltz et al. 1994). This enzyme could possibly have a similar function in sorghum, another C_4 plant.

5.5 Hydrolytic Enzymes

The hydroxynitrile lyase (HNL) spots (20 and 28; Table 7.1) were classified as hydrolytic enzyme (7.1%). These enzymes catalyse the cleavage of cyanogenic glycosides into aldehydes or ketones and hydrogen cyanide (Lauble et al. 2002;

Purkarthofer et al. 2007). Hydrogen cyanide is toxic, and its production (cyanogenesis) is initiated by tissue damage (White et al. 1998). In plants, cyanogenesis acts as a defence mechanism against herbivores and microbial attack as well as alternative nitrogen source for amino acid synthesis by young seedlings (Fechter and Griengl 2004; Hickel et al. 1996). In all cyanogenic plants, such as sorghum and cassava, cyanogenic glucosides (substrate) and the corresponding enzymes are located in different cellular compartments (Morant et al. 2008). This spatial separation of the substrate and enzyme helps prevent cyanogenesis until the tissue is damaged (Poulton 1990; Vetter 2000). In sorghum, cyanogenic glucosides are located in the vacuoles of epidermal cells while the HNL are located in the cytoplasm of mesophyll cells (Hickel et al. 1996).

5.6 Detoxifying Enzymes

Glutathione *S* transferase (GST; EC 2.5.1.18) I (spot 26; Table 7.1) was identified in our sorghum leaf proteome. Plant GSTs are mainly cytosolic enzymes (Dixon et al. 2002), occurring either as homo- or heterodimers, each with an MW range of between 25 and 27 kDa (Edwards et al. 2000). Therefore, it is highly probable that the 26 kDa GST protein (spot 26; Table 7.1) identified in this study could represent a subunit of the enzyme. Functionally, GSTs are glutathione-dependent detoxifying enzymes, which conjugate glutathione to a wide range of natural products, environmental toxins (such as herbicides) as well as products of oxidative stress. The glutathione conjugates are then transported to the vacuoles for further metabolism into a range of sulphur-containing metabolites (Edwards et al. 2000). The observation of a GST protein spot in our sorghum leaf proteome provided evidence that these enzymes are present in plant cells even under non-stress conditions. However, their precise functions in normal cellular processes are still not clearly understood. It is proposed that GSTs may have a function in detoxifying endogenous products of oxidative damage such as membrane lipid peroxides and products of oxidative DNA damage. GSTs might also act as nonenzymatic carrier proteins (ligandins), which bind and transport plant hormones such as indoleacetic acid (Marrs 1996).

6 Subcellular Localisation of Identified Sorghum Leaf Proteins

Knowledge of protein localisation into subcellular compartments is important as it helps clarify protein function and mechanisms of action (Kumar et al. 2002; van Wijk 2004). Subcellular localisations of the identified sorghum leaf proteins were predicted using a combination of TargetP version 1.1 (http://www.cbs.dtu.dk/services/TargetP/) (Emanuelsson et al. 2007), Predotar version 1.0 (http://urgi.versailles.inra.fr/predotar/predotar.html) (Small et al. 2004) and literature sources.

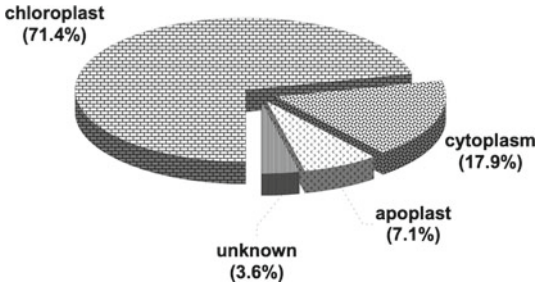

Fig. 7.6 Subcellular localisation results of the identified sorghum leaf proteins. Subcellular localisations of the sorghum leaf proteins were predicted using a combination of TargetP version 1.1 (http://www.cbs.dtu.dk/services/TargetP/; Emanuelsson et al. 2007), Predotar version 1.03 (http://urgi.versailles.inra.fr/predotar/predotar.html; Small et al. 2004) and literature sources. *Numbers* indicated in *brackets* represent the proportion of proteins within each subcellular compartment expressed as a percentage of the 28 spots that were positively identified by a combination of MALDI-TOF and MALDI-TOF–TOF MS and database searches

The subcellular localisation of each of the positively identified proteins is given in Table 7.1 while a graphical representation of the total number of proteins in each subcellular location is shown in Fig. 7.6.

Sorghum leaf proteins identified in this study as outlined in Sect. 5 above were predicted to be localised in the chloroplast (20 spots; 71.4%), followed by the cytoplasm (5 spots; 17.9%), apoplast (2 spots; 7.1%) and an unknown location (1 spot; 3.6%). These results are consistent with the observation that the most prominent functional category amongst the identified sorghum leaf proteins (Fig. 7.4; Table 7.1) was carbohydrate metabolism and photosynthesis-related processes (Fig. 7.5; Table 7.1). In green plants such as sorghum, photosynthesis and carbohydrate metabolic pathways occur primarily in photosynthetic organelles called chloroplasts (van Wijk 2004). Therefore, the dominance of the chloroplast as the main subcellular location of sorghum leaf proteins in the current study correlates well with the functional classification results (Table 7.1; Fig. 7.5).

The chloroplast itself has different compartments: (1) the outer and inner envelop membranes that surround the organelle, (2) the soluble stroma, (3) the thylakoid membrane and (4) the thylakoid lumen (van Wijk 2004). Each of these compartments has a different subset of proteins or subproteomes. Collectively, the chloroplast has a diverse population of proteins, which are either soluble or membrane associated. The representation of these two broad classes of proteins in any proteomic study also differs depending on the extraction procedures utilised in the experiment. Membrane proteins are usually poorly represented because of their hydrophobic nature and thus low solubility in most 2DE extraction and solubilisation buffers (Molloy 2000; Santoni et al. 2000). To determine the sub-organellar locations of the 20-chloroplast predicted proteins (Table 7.1), bioinformatics searches were carried out on the Plant Proteome Database (PPDB; http://ppdb.tc.cornell.edu/) (Sun et al. 2009) using *A. thaliana* homologs. Table 7.2 gives a

Table 7.2 List of stromal and membrane-associated sorghum leaf chloroplast proteins

			Membrane	
Protein name	Spot[a]	Stroma	Integral	Peripheral
Pyruvate phosphate dikinase	1 and 2	+		
RuBisCo	5 and 6	+		
Chloroplastic NADP-dependent malate dehydrogenase	10	+		
Sedoheptulose-1,7-biphosphate precursor	11	+		
Phosphoribulikinase, chloroplastic precursor	12	+		
Ferrodoxin-NADP oxidoreductase	17 and 18			+
Fructose 1,6-biphosphate aldolase precursor	34	+		
Oxygen-evolving enhancer protein 1 precursor	35			+
PSII oxygen-evolving complex protein 2	36			+
Chlorophyll *a*/*b*-binding proteins type III precursor	37		+	
ATP synthase	7, 32, 33, 13 and 40			+
Translational elongation factor Tu	8	+		
Putatative adenylate kinase, chloroplast	24	+		

[a]Spot number as indicated on the 2DE gel image (Fig. 7.4)
+ Indicates subcellular localisation of proteins

summary of the chloroplast subproteome protein identities and their locations (stroma versus membrane). The membrane proteins were further divided into peripheral or integral proteins depending on their interaction with the membrane (Friso et al. 2004). From the results, it was observed that both the stromal and membrane proteins were represented in equal proportions. Of the ten membrane-associated proteins, only one, the chlorophyll *a*/*b* binding protein (spot 37; Table 7.2), was an integral protein while the other nine were associated with either the lumenal or the stromal side of the thylakoid membrane.

7 Conclusion

All cells of an organism have the same genome; however the proteomes of different cells, tissues and organs within the same organism differ in accordance with their functions. As shown in mammalian systems, proteomics studies increase the understanding of molecular mechanisms underlying reproduction, growth and development of these organisms. Plants are indispensable to most mammals, including humans—as source of food, fuel and shelter material, contributing the largest biomass on earth. Despite this, the development of plant proteomics is much slower, underfunded, understudied and therefore poorly understood in comparison to animal studies. Although some extrapolations can be made, plants and animals are very

different and in order to understand plant molecular biology better, we need to know their proteomes and factors that affect them in as much detail as possible. Few plants are currently receiving proteomics attention, with sorghum, and sugarcane to some extent, being the only representative of the Saccharinae to our knowledge. As in mammalian and prokaryotic systems, through proteomics, researchers will identify and characterise protein biomarkers for various aspects related to Saccharinae genetics, phenotypic features, reproduction, growth and development offering novel opportunities for productivity improvement, disease prevention and/or treatment.

Acknowledgments We would like to thank Professor Antoni Slabas and Dr. Bill Simon of Durham University for the MS–MS protein identification data, the National Research Foundation of South Africa (NRF) and the South African Department of Science and Technology (DST) for research funds.

References

Aebersold R, Goodlett DR (2001) Mass spectrometry in proteomics. Chem Rev 101:269–295

Albertin W, Langella O, Joets J, Negroni L, Zivy M, Damerval C, Thiellement H (2009) Comparative proteomics of leaf, stem, and root tissues of synthetic *Brassica napus*. Proteomics 9:793–799

Andersson I, Backlund A (2008) Structure and function of Rubisco. Plant Physiol Biochem 46:275–291

Arakaki AK, Ceccarelli EA, Carrillo N (1997) Plant-type ferredoxin-NADP+reductases: a basal structural framework and a multiplicity of functions. Fed Am Soc Exp Biol J 11:133–140

Bannai H, Tamada Y, Maruyama O, Nakai K, Miyano S (2002) Extensive feature detection of N-terminal protein sorting signals. Bioinforma 18:298–305

Bassi R, Sandona D, Croe R (1997) Novel aspects of chlorophyll a/b-binding proteins. Physiol Plant 100:769–779

Blackstock WP, Weir MP (1999) Proteomics: quantitative and physical mapping of cellular proteins. Trends Biotechnol 17:121–127

Boehlein SK, Sewell AK, Cross J, Stewart JD, Hannah LC (2005) Purification and characterization of adenosine diphosphate glucose pyrophosphorylase from maize/potato mosaics. Plant Physiol 138:1552–1562

Bradford MM (1976) A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248–254

Carpentier SC, Coemans B, Podevin N, Laukens K, Witters E, Matsumura H, Terauchi R, Swennen R, Panis B (2008) Functional genomics in a non-model crop: transcriptomics or proteomics? Physiol Plant 133:117–130

Carrari F, Coll-Garcia D, Schauer N, Lytovchenko A, Palacios-Rojas N, Balbo I, Rosso M, Fernie AR (2005) Deficiency of a plastidial adenylate kinase in Arabidopsis results in elevated photosynthetic amino acid biosynthesis and enhanced growth. Plant Physiol 137:70–82

Claros MG, Vincens P (1996) Computational method to predict mitochondrially imported proteins and their targeting sequences. Eur J Biochem 241:779–786

Cokol M, Nair R, Rost B (2000) Finding nuclear localization signals. Eur Mol Biol Organ Rep 1:411–415

Cvetic T, Veljovic-Jovanovic S, Vucinic Z (2008) Characterization of NAD-dependent malate dehydrogenases from spinach leaves. Protoplasma 232:247–253

Ding Y, Ma QH (2004) Characterization of a cytosolic malate dehydrogenase cDNA which encodes an isozyme toward oxaloacetate reduction in wheat. Biochimie 86:509–518

Dixon DP, Lapthorn A, Edwards R (2002) Plant glutathione transferases. Genome Biol 3:3
Dunn MJ, Gorg A (2001) Two-dimensional polyacrylamide gel electrophoresis for proteome analysis. In: Pennington SR, Dunn MJ (eds) Proteomics from protein sequence to function. BIOS Scientific Publishers, Oxford, pp 43–63
Edwards R, Dixon DP, Walbot V (2000) Plant glutathione S-transferases: enzymes with multiple functions in sickness and in health. Trends Plant Sci 5:193–198
Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc 2:953–971
Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol 300:1005–1016
Fechter MH, Griengl H (2004) Hydroxynitrile lyases: biological sources and application as biocatalysts. Food Technol Biotechnol 42:287–294
Fridlyand LE, Backhausen JE, Scheibe R (1998) Flux control of the malate valve in leaf cells. Arch Biochem Biophys 349:290–298
Friso G, Giacomelli L, Ytterberg AJ, Peltier JB, Rudella A, Sun Q, van Wijk KJ (2004) In-depth analysis of the thylakoid membrane proteome of *Arabidopsis thaliana* chloroplasts: new proteins, new functions, and a plastid proteome database. Plant Cell 16:478–499
Ganeteg U, Strand A, Gustafsson P, Jansson S (2001) The properties of the chlorophyll a/b-binding proteins Lhca2 and Lhca3 studied in vivo using antisense inhibition. Plant Physiol 127:150–158
Gorg A, Obermaier C, Boguth G, Harder A, Scheibe B, Wildgruber R, Weiss W (2000) The current state of two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis 21: 1037–1053
Gorg A, Weiss W (2004) Protein profile comparisons of microorganisms, cells and tissues using 2D gels. In: Speicher DW (ed) Proteome analysis: interpreting the genome. Elsevier, New York, pp 19–73
Goward CR, Nicholls DJ (1994) Malate dehydrogenase: a model for structure, evolution, and catalysis. Protein Sci 3:1883–1888
Guan HP, Keeling PL (1998) Starch Biosynthesis: understanding the functions and interactions of multiple isoenzymes of starch synthase and branching enzyme. Trends Glycosci Glycotechnol 10:307–319
Gummadova JO, Fletcher GJ, Moolna A, Hanke GT, Hase T, Bowsher CG (2007) Expression of multiple forms of ferredoxin NADP+oxidoreductase in wheat leaves. J Exp Bot 58:3971–3985
Harris EH, Boynton JE, Gillham NW (1994) Chloroplast ribosomes and protein synthesis. Microbiol Rev 58:700–754
Heazlewood JL, Millar AH (2003) Integrated plant proteomics - putting the green genomes to work. Funct Plant Biol 30:471–482
Heide H, Kalisz HM, Follmann H (2004) The oxygen evolving enhancer protein 1 (OEE) of photosystem II in green algae exhibits thioredoxin activity. J Plant Physiol 161:139–149
Heldt HW (1997) Plant biochemistry and molecular biology. Oxford University Press, Oxford
Hickel A, Hasslacher M, Griengl H (1996) Hydroxynitrile lyases: functions and properties. Physiol Plant 98:891–898
Hua S, Sun Z (2001) Support vector machine approach for protein subcellular localization prediction. Bioinformatics 17:721–728
Hurkman WJ, Tanaka CK (2007) High-resolution two-dimensional gel electrophoresis: a cornerstone for plant proteomics. In: Samaj J, Thelen JJ (eds) Plant proteomics. Springer, Berlin, pp 14–28
Ifuku K, Nakatsu T, Shimamoto R, Yamamoto Y, Ishihara S, Kato H, Sato F (2005) Structure and function of the PsbP protein of photosystem II from higher plants. Photosyn Res 84:251–255
Igamberdiev AU, Kleczkowski LA (2006) Equilibration of adenylates in the mitochondrial intermembrane space maintains respiration and regulates cytosolic metabolism. J Exp Bot 57: 2133–2141
International rice genome sequencing project (2005) The map-based sequence of the rice genome. Nature 436:793–800

Jangpromma N, Kitthaisong S, Daduang S, Jaisil P, Thammasirirak S (2007) 18 kDa protein accumulation in sugarcane leaves under drought stress conditions. KMITL Sci Technol J 7:44–54
Jorrin JV, Maldonado AM, Castillejo MA (2007) Plant proteome analysis: a 2006 update. Proteomics 7:2947–2962
Jorrin-Novo JV, Maldonado AM, Echevarria-Zomeno S, Valledor L, Castillejo MA, Curto M, Valero J, Sghaier B, Donoso G, Redondo I (2009) Plant proteomics update (2007-2008): second-generation proteomic techniques, an appropriate experimental design, and data analysis to fulfill MIAPE standards, increase plant proteome coverage and expand biological knowledge. J Proteomics 72:285–314
Kang IH, Lee JW, Lee JH, Kang CJ, Sim W-S, Kim J-K (1998) Light-independent regulation of chloroplast translational elongation factor Tu gene expression in three types of grass: rice, maize, and barley. J Plant Biol 41:324–329
Kellogg EA, Juliano ND (1997) The structure and function of RuBisCo and their implications for systematic studies. Am J Bot 84:413–428
Komatsu S (2006) Plant proteomics databases: their status in 2005. Curr Bioinform 1:33–36
Krishnamurthy L, Serraj R, Hash CT, Dakheel AJ, Reddy BVS (2007) Screening sorghum genotypes for salinity tolerant biomass production. Euphytica 156:15–24
Kruger NJ (1997) Carbohydrate synthesis and degradation. In: Dennis DT, Turpin DH, Lefebvre DD, Layzell DB (eds) Plant metabolism, 2nd edn. Addison Wesley Longman, Essex, pp 83–104
Kumar A, Agarwal S, Heyman JA, Matson S, Heidtman M, Piccirillo S, Umansky L, Drawid A, Jansen R, Liu Y, Cheung KH, Miller P, Gerstein M, Roeder GS, Snyder M (2002) Subcellular localization of the yeast proteome. Genes Dev 16:707–719
Lange PR, Geserick C, Tischendorf G, Zrenner R (2008) Functions of chloroplastic adenylate kinases in Arabidopsis. Plant Physiol 146:492–504
Lauble H, Miehlich B, Forster S, Wajant H, Effenberger F (2002) Crystal structure of hydroxynitrile lyase from *Sorghum bicolor* in complex with the inhibitor benzoic acid: a novel cyanogenic enzyme. Biochemie 41:12043–12050
Liebler DC (ed) (2004) Proteomics in cancer research. Wiley, New Jersey
Lin D, Tabb DL, Yates JR 3rd (2003) Large-scale protein identification using mass spectrometry. Biochim Biophys 1646:1–10
Macdonald FD, Buchanan BB (1997) The reductive pentose phosphate pathway and its regulation. In: Dennis DT, Turpin DH, Lefebvre DD, Layzell DB (eds) Plant metabolism, 2nd edn. Addison Wesley Longman, Essex, pp 299–313
Marengo E, Robotti E, Antonucci F, Cecconi D, Campostrini N, Righetti PG (2005) Numerical approaches for quantitative analysis of two-dimensional maps: a review of commercial software and home-made systems. Proteomics 5:654–666
Marrs KA (1996) The functions and regulation of glutathione S-transferases in plants. Annu Rev Plant Physiol Plant Mol Biol 47:127–158
Martin C, Smith AM (1995) Starch biosynthesis. Plant Cell 7:971–985
Matthiesen R, Mutenda KE (2007) Introduction to proteomics. Methods Mol Biol 367:1–35
McCarty RE (1992) A plant biochemist's view of H+-ATPases and ATP synthesases. J Exp Bot 172:431–441
McEvoy JP, Brudvig GW (2006) Water-splitting chemistry of photosystem II. Chem Rev 106:4455–4483
Minarik P, Tomaskova N, Kollarova M, Antalik M (2002) Malate dehydrogenases—structure and function. Gen Physiol Biophys 21:257–265
Molloy MP (2000) Two-dimensional electrophoresis of membrane proteins using immobilized pH gradients. Anal Biochem 280:1–10
Morant AV, Jorgensen K, Jorgensen C, Paquette SM, Sanchez-Perez R, Moller BL, Bak S (2008) Beta-glucosidases as detonators of plant chemical defense. Phytochemistry 69:1795–1813
Murashige T, Skoog F (1962) A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol Plant 15:473–497
Nakai K, Horton P (1999) PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. Trends Biochem Sci 24:34–36

Ndimba BK, Chivasa S, Hamilton JM, Simon WJ, Slabas AR (2003) Proteomic analysis of changes in the extracellular matrix of Arabidopsis cell suspension cultures induced by fungal elicitors. Proteomics 3:1047–1059

Ndimba BK, Thomas LA (2008) Proteomics in South Africa: current status, challenges and prospects. Biotechnol J 3:1368–1374

Ng JH, Ilag LL (2002) Functional proteomics: separating the substance from the hype. Drug Discov Today 7:504–505

Ngara R (2009) A Proteomic analysis of drought and salt stress responsive proteins of different sorghum varieties. Biotechnology. University of the Western Cape, Cape Town, p 346

Ngara R, Rees J, Ndimba BK (2008) Establishment of sorghum cell suspension culture system for proteomics studies. Afr J Biotechnol 7:744–749

Nozu Y, Tsugita A, Kamijo K (2006) Proteomic analysis of rice leaf, stem and root tissues during growth course. Proteomics 6:3665–3670

Pandey A, Mann M (2000) Proteomics to study genes and genomes. Nature 405:837–846

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob-ur R, Ware D, Westhoff P, Mayer KF, Messing J, Rokhsar DS (2009) The Sorghum bicolor genome and the diversification of grasses. Nature 457:551–556

Patterson SD (2000) Mass spectrometry and proteomics. Physiol Genomics 2:59–65

Patterson SD (2004) How much of the proteome do we see with discovery-based proteomics methods and how much do we need to see? Curr Proteomics 1:3–12

Patterson SD, Aebersold R, Goodlett DR (2001) Mass spectrometry-based methods for protein identification and phosphorylation site analysis. In: Pennington SR, Dunn MJ (eds) Proteomics from protein sequence to function. BIOS Scientific Publishers, Oxford, pp 87–130

Patterson SD, Aebersold RH (2003) Proteomics: the first decade and beyond. Nat Genet 33(Suppl):311–323

Patton WF (2000) A thousand points of light: the application of fluorescence detection technologies to two-dimensional gel electrophoresis and proteomics. Electrophoresis 21:1123–1144

Podolak E (2010) Sequencing's new race. Biotechniques 48:105–111

Porubleva L, Vander Velden K, Kothari S, Oliver DJ, Chitnis PR (2001) The proteome of maize leaves: use of gene sequences and expressed sequence tag data for identification of proteins with peptide mass fingerprints. Electrophoresis 22:1724–1738

Poulton JE (1990) Cyanogenesis in plants. Plant Physiol 94:401–405

Preiss J (1997) Modulation of starch synthesis. In: Foyer CH, Quick WP (eds) A molecular approach to primary metabolism in higher plants. Taylor and Francis Publishers, London, pp 81–104

Purkarthofer T, Skranc W, Schuster C, Griengl H (2007) Potential and capabilities of hydroxynitrile lyases as biocatalysts in the chemical industry. Appl Microbiol Biotechnol 76:309–320

Rabilloud T (2002) Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains. Proteomics 2:3–10

Raines CA (2003) The Calvin cycle revisited. Photosyn Res 75:1–10

Raymond J, Blankenship RE (2008) The origin of the oxygen-evolving complex. Coord Chem Rev 252:377–383

Salekdeh GH, Komatsu S (2007) Crop proteomics: aim at sustainable agriculture of tomorrow. Proteomics 7:2976–2996

Santoni V, Kieffer S, Desclaux D, Masson F, Rabilloud T (2000) Membrane proteomics: use of additive main effects with multiplicative interaction model to classify plasma membrane proteins according to their solubility and electrophoretic properties. Electrophoresis 21:3329–3344

Scheibe R (2004) Malate valves to balance cellular energy supply. Physiol Plant 120:21–26

Schiltz E, Burger S, Grafmuller R, Deppert WR, Haehnel W, Wagner E (1994) Primary structure of maize chloroplast adenylate kinase. Eur J Biochem 222:949–954

Schiltz S, Gallardo K, Huart M, Negroni L, Sommerer N, Burstin J (2004) Proteome reference maps of vegetative tissues in pea. An investigation of nitrogen mobilization from leaves during seed filling. Plant Physiol 135:2241–2260

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen W, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He R, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin J, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren L, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Van Buren P, Vaughn MW, Ying K, Yeh CT, Emrich SJ, Jia Y, Kalyanaraman A, Hsia AP, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia JM, Deragon JM, Estill JC, Fu Y, Jeddeloh JA, Han Y, Lee H, Li P, Lisch DR, Liu S, Liu Z, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang L, Yu Y, Zhang L, Zhou S, Zhu Q, Bennetzen JL, Dawe RK, Jiang J, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA, Wilson RK (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115

Small I, Peeters N, Legeai F, Lurin C (2004) Predotar: a tool for rapidly screening proteomes for N-terminal targeting sequences. Proteomics 4:1581–1590

Speicher DW (2004) Overview of proteome analysis. In: Speicher DW (ed) Proteome analysis: interpreting the genome. Elsevier, New York, pp 1–18

Sproviero EM, Gascon JA, McEvoy JP, Brudvig GW, Batista VS (2007) Quantum mechanics/molecular mechanics structural models of the oxygen-evolving complex of photosystem II. Curr Opin Struct Biol 17:173–180

Sun Q, Zybailov B, Majeran W, Friso G, Olinares PD, van Wijk KJ (2009) PPDB, the plant proteomics database at Cornell. Nucleic Acids Res 37:D969–D974

Tabita FR, Hanson TE, Li H, Satagopan S, Singh J, Chan S (2007) Function, structure, and evolution of the RubisCO-like proteins and their RubisCO homologs. Microbiol Mol Biol Rev 71:576–599

Tamoi M, Nagaoka M, Yabuta Y, Shigeoka S (2005) Carbon metabolism in the Calvin cycle. Plant Biotechnol 22:355–360

Tetlow IJ, Davies EJ, Vardy KA, Bowsher CG, Burrell MM, Emes MJ (2003) Subcellular localization of ADP glucose pyrophosphorylase in developing wheat endosperm and analysis of the properties of a plastidial isoform. J Exp Bot 54:715–725

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408:796–815

Thiellement H, Bahrman N, Damerval C, Plomion C, Rossignol M, Santoni V, de Vienne D, Zivy M (1999) Proteomics for genetic and physiological studies in plants. Electrophoresis 20:2013–2026

Thomas JC, Ughy B, Lagoutte B, Ajlani G (2006) A second isoform of the ferredoxin:NADP oxidoreductase generated by an in-frame initiation of translation. Proc Natl Acad Sci U S A 103:18368–18373

van Wijk KJ (2001) Challenges and prospects of plant proteomics. Plant Physiol 126:501–508

van Wijk KJ (2004) Plastid proteomics. Plant Physiol Biochem 42:963–977

Vetter J (2000) Plant cyanogenic glucosides. Toxicon 38:11–36

Vettore AL, da Silva FR, Kemper EL, Souza GM, da Silva AM, Ferro MI, Henrique-Silva F, Giglioti EA, Lemos MV, Coutinho LL, Nobrega MP, Carrer H, Franca SC, Bacci Junior M, Goldman MH, Gomes SL, Nunes LR, Camargo LE, Siqueira WJ, Van Sluys MA, Thiemann

OH, Kuramae EE, Santelli RV, Marino CL, Targon ML, Ferro JA, Silveira HC, Marini DC, Lemos EG, Monteiro-Vitorello CB, Tambor JH, Carraro DM, Roberto PG, Martins VG, Goldman GH, de Oliveira RC, Truffi D, Colombo CA, Rossi M, de Araujo PG, Sculaccio SA, Angella A, Lima MM, de Rosa Junior VE, Siviero F, Coscrato VE, Machado MA, Grivet L, Di Mauro SM, Nobrega FG, Menck CF, Braga MD, Telles GP, Cara FA, Pedrosa G, Meidanis J, Arruda P (2003) Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. Genome Res 13:2725–2735

Vicentini R, Menossi M (2009) The predicted subcellular localisation of the sugarcane proteome. Funct Plant Biol 36:242–250

von Ballmoos C, Dimroth P (2007) Two distinct proton binding sites in the ATP synthase family. Biochemistry 46:11800–11809

Wajant H, Mundry K-W (1993) Hydroxynitrile lyase from S*orghum bicolor*: a glycoprotein heterotetramer. Plant Sci 89:127–133

Washburn MP, Wolters D, Yates JR III (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol 19:242–247

Watson BS, Asirvatham VS, Wang L, Sumner LW (2003) Mapping the proteome of barrel medic (*Medicago truncatula*). Plant Physiol 131:1104–1123

Westermeier R (2005) Electrophoresis in practice, 4th edn. Wiley-VCH, Weinheim

White WLB, Arias-Garzon DI, McMahon JM, Sayre RT (1998) Cyanogenesis in cassava- the role of hydroxynitrile lyase in root cyanide production. Plant Physiol 116:1219–1225

Chapter 8
Gene Mutagenesis Systems and Resources for the Saccharinae

Zhanguo Xin, Ming-Li Wang, Surinder Chopra, and Pohao Wang

Abstract This chapter focuses on mutant populations in Saccharinae that are available for genomic studies. Emphasis is on sorghum mutant resources as few mutant resources are available in sugarcane or Miscanthus due to polyploidy of their genomes. As a minimally redundant genome that last experienced genome duplication ~70 million years ago, sorghum is particularly sensitive to ethyl methane sulfonate (EMS) and other mutagens, with many mutagenized lines displaying various phenotypes at EMS concentrations as low as 0.1%. Many mutant phenotypes have the potential to increase biomass production or bioenergy conversion efficiency of sorghum plants. Characterizing these sorghum mutants may also provide useful information to improve biomass production and bioenergy conversion efficiency or other traits of other plants, for example via RNAi technology. The small size of the sorghum genome, its diploid nature, and high gene density also make sorghum an obvious choice for an efficient transposon tagging system. The *Candystripe1* (*Cs1*) transposon has been isolated. The activity of *Cs1* has been demonstrated in the *y1* locus, and several mutations have been isolated. An understanding of the genetic behavior of this element has been gathered toward the development of a viable transposon tagging system in sorghum.

Keywords Sorghum • Mutagenesis • Transposon • Ethyl methane sulfonate • TILLING • *Bmr* • *Erl*

Z. Xin (✉)
Plant Stress and Germplasm Development Unit, USDA-ARS,
3810 4th Street, Lubbock, TX 79415, USA
e-mail: zhanguo.xin@ars.usda.gov

M.-L. Wang
PGRCU, USDA-ARS, 1109 Experiment Street, Griffin, GA 30223, USA

S. Chopra • P. Wang
Crop and Soil Science Department, Pennsylvania State University,
University Park, PA 16802, USA

A.H. Paterson (ed.), *Genomics of the Saccharinae*, Plant Genetics and Genomics: Crops and Models 11, DOI 10.1007/978-1-4419-5947-8_8,

1 Introduction

Mutations, either natural or induced, provide raw material for evolution and plant breeding. Using modern molecular genetics techniques, mutational analyses play a crucial role in elucidation of gene function, metabolic processes, signaling, growth, and development. Mutational analyses fall into two broad categories: forward and reverse genetics. In forward genetics, informative mutant phenotypes are identified first. Then, the mutant with the desired trait is crossed to another ecotype that has extensive DNA polymorphism and the genomic region attributable to the phenotype is identified through genetic mapping. The mutated gene responsible for the phenotype is identified through traditional positional cloning or through sequencing candidate genes if sufficient information is available to infer the processes leading to the phenotype. In reverse genetics, a series of mutants for a gene is first isolated based on sequence differences from the wild-type sequence. Then, the mutant phenotypes are analyzed to deduce the function of the gene. Reverse genetics techniques are now available to enable high-throughput analysis of gene function on a genome-wide scale. Sorghum genome sequencing from a leading inbred BTx623 has been completed (Paterson et al. 2009b). The focus now is to establish the function of majority of the genes in the sorghum genome. A well-characterized mutant population and reverse genetics techniques will likely play an important role in establishing the function of genes, especially those for which a functional assay is not available.

2 Annotated Individually Pedigreed Mutagenized Sorghum Library

Mutagenesis has long been applied to sorghum to isolate novel phenotypes that may have potential application in breeding (Gaul 1964; Quinby and Karper 1942). Many mutants with unique phenotypes that have not been observed in natural sorghum collections have been selected from populations treated with various mutagens, such as X-ray and γ-irradiation, ethyl methane sulfonate (EMS), methyl methane sulfonate (MMS), diethyl sulfate (DES), *N*-nitroso methyl urea (NMU), *N*-nitroso ethyl urea (NEU), or combinations of chemical and irradiation mutagens (Quinby and Karper 1942; Sree Ramulu 1970a, b; Sree Ramulu and Sree Rangasamy 1972). Many beneficial mutations, including dwarfing, early flowering, high protein digestibility, high lysine, and others, have been widely used in sorghum breeding (Ejeta and Axtell 1985; Oria et al. 2000; Quinby 1975; Singh and Axtell 1973). The late Dr. Keith Schertz, a former sorghum geneticist with USDA-ARS, collected and preserved more than 400 natural and induced mutant lines from various genetic backgrounds. This collection, now under the curation of the Plant Stress and Germplasm Development Unit (USDA-ARS, Lubbock, TX), provides a starting point to study the functions of sorghum genes.

2.1 Development of an Annotated Individually Pedigreed Mutagenized Sorghum Library

The sorghum genome sequence and the identification of its genes have made it possible to study gene function on a genome-wide scale, and to compare gene function with other plants (Paterson 2008; Paterson et al. 2009b). A systematic mutant library that contains multiple mutations for all genes in the sorghum genome is urgently needed to deduce the functions of sorghum genes. Xin et al. (2008) reported a modest population of 768 pedigreed EMS-mutagenized lines of BTx623, a leading inbred used for sorghum genome sequencing. The mutant library was developed by single-seed-descent from individual mutagenized seeds (M_1) to M_3 generation. Genomic DNA was prepared with leaf samples collected from the M_2 plants used to produce M_3 seeds. Phenotypes are annotated at the M_3 generation to ensure that any phenotype observed in a family is descended from a single mutagenesis event (a single germ cell), represented by an M_2 plant used to prepare genome DNA. Following phenotype annotation, ten M_3 panicles are bulked as M_4 seeds, which are deposited in the library and will be distributed to end users on request. The mutant library is named the *A*nnotated *I*ndividually pedigreed *M*utagenized *S*orghum (AIMS) library. Since the mutant library is pedigreed, recessive lethal mutations can be preserved in heterozygous state. A pilot study shows that the library has a mutation rate about 1/526 kb (Xin et al. 2008). Given the ~730 Mb genome size of sorghum and the finding that about ¼ of the DNA is euchromatin (Paterson et al. 2009b), each mutant line is expected to harbor about 340 mutations in the euchromatin. A mutant library with 6,400 lines would contain more than two million independent mutations, i.e., about 80 mutations per gene. This level of coverage, although far from saturation mutagenesis, should provide an adequate resource for genome-wide identification of mutant series for most genes in the genome and to screen for mutants that can be used for sorghum improvement or biological studies. The library has now expanded to over 5,000 lines and will be expanded to 6,400 lines in the next 2 years. The mutant library can be accessed online at http://www.lbk.ars.usda.gov/psgd/index-sorghum.aspx.

Many factors affect the quality of mutant libraries. The first important factor is the choice of mutagens. EMS was used to generate the AIMS library because of its high rate of success in sorghum and many other plant species (Greene et al. 2003). In a comparative study of multiple mutagens, EMS is shown to induce ten times more chlorophyll mutations than NUE and MMS (Sree Ramulu 1970b). It has been used extensively to create sorghum mutants with useful traits such as early flowering, dwarfing, and a series of mutants with no or sparse epicuticular wax layers (Jenks et al. 1994; Peters et al. 2009; Singh and Drolsom 1974; Sree Ramulu 1970b). The second factor is the dosage of the mutagen used. The concentration of EMS used to generate the mutant library must be evaluated carefully to balance seed setting with adequate mutation frequency. This will involve trial and error and may vary for different varieties or even different batches of EMS (Henikoff and Comai 2003). BTx623 is very sensitive to EMS treatment. At 0.1% (v/v) EMS, only 40% of M_1

plants set seeds (Xin et al. 2008). This concentration is much lower than the 0.3% (v/v) EMS which is frequently used in *Arabidopsis* and many other organisms (Greene et al. 2003). The highest concentration of EMS that can be tolerated by BTx623 is 0.25%, at which less than 10% of the plants produced seeds (Xin et al. 2008). Thus, the mutant library is generated with a series of EMS concentration ranging from 0.1 to 0.25%, to balance mutation frequency with survival of mutants. Other factors also impact the establishment of useful TILLING populations in sorghum. For example, cross-pollination must be vigorously controlled to produce a high-quality mutant library. Under normal growth conditions, sorghum is predominantly self-fertilized with a cross-fertilization rate ranging from 5 to 10% (Ellstrand and Foster 1983). After EMS-mutagenesis, cross-fertilization increased dramatically. A previous sorghum mutagenesis attempt was unsuccessful when cloth bags (Lawson Bags, Northfield, IL) failed to prevent cross-pollination. An examination of resulting M_2 plants using four hyperpolymorphic sorghum simple sequence repeat (SSR) markers, Xtxp287, Xtxp270, Xtxp51, and Xtxp295 [publicly available (Menz et al. 2002)], showed that over 30% of the M_2 plants were the result of cross-pollination from unknown sources. Cross-pollination can be effectively controlled by covering the panicles at each generation with rainproof paper pollination bags (Lawson Bags, Northfield, IL) before anthesis. Corn earworms and birds also pose serious threats to the limited seed set in M_1 plants during the grain-filling period. The pollination bags must be injected with pesticide to control corn earworm. Despite these challenges, a sizable mutant library has been established and ready for distribution.

Ongoing phenotype annotation shows that this mutant library displays a variety of phenotypes, potentially serving as both a forward genetic resource for identifying useful traits and their genes for sorghum improvement, and as a reverse genetic resource for identifying mutant series in specific genes to deduce their functions. A selection of phenotypes is presented in Fig. 8.1. Here, we discuss two traits that may be useful for improving conversion efficiency of sorghum stover to ethanol and biomass production. Readers are referred to the online database for a complete compilation of phenotypes observed in the mutant library.

2.2 Brown Midrib Mutations

Brown midrib (*bmr*) mutants have been isolated from C4 cereals such as maize, sorghum, and millet through natural or induced mutations (Sattler et al. 2010). The mutant phenotype is typified by a distinctive brownish colored mid veins of leaves, which can be easily identified in the field. A typical *bmr* mutant is shown in Fig. 8.1. Some mutants also accumulate reddish brown to yellow pigment in the stalk, root, and stem pith. The *bmr* mutation is associated with reduced lignin content, increased digestibility for livestock, and increased conversion efficiency of sorghum stover to ethanol (Vermerris et al. 2007).

Sorghum *bmr* mutants were first isolated by Porter et al. (1978) from diethyl sulfate mutagenized population. Twenty-eight sorghum *bmr* mutants represented by

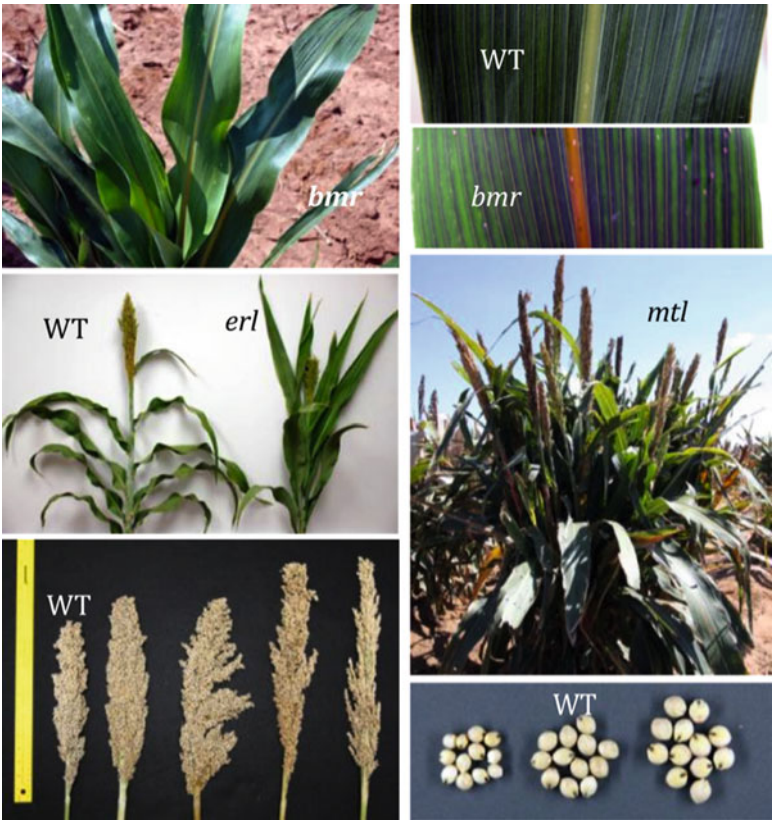

Fig. 8.1 A selection of mutant phenotypes that may have potential to improve bioenergy conversion efficiency and biomass production in sorghum. A complete collection of mutant phenotypes can be found online (http://www.lbk.ars.usda.gov/psgd/index-sorghum.aspx)

four loci (*bmr2*, *bmr6*, *bmr12*, and *bmr19*) have been isolated from various sources including natural mutation (Sattler et al. 2010). Two of the loci have been cloned by candidate gene approaches. The *bmr6* mutation encodes a cinnamyl alcohol dehydrogenase (CAD) and *bmr12* encodes a caffeic *O*-methyl transferase (COMT) (Bout and Vermerris 2003; Saballos et al. 2009; Sattler et al. 2009). These two enzymes are involved in the last two steps of biosynthesis of monolignols, the precursors for lignin biosynthesis. Among these four loci, *bmr2* and *bmr19* are represented by a single locus, indicating that saturation mutagenesis has not been achieved (Sattler et al. 2010). Moreover, both *bmr6* and *bmr12*, which are the main sources for commercial *bmr* forage sorghum, are complete knockout mutation. To identify additional *bmr* mutants and to isolate non-knockout alleles of *bmr6* and *bmr12*, we initiated a systematic approach to isolate additional *bmr* mutants. A close inspection of about 3,000 individual M_3 families identified over 100 independent mutants with the typical brownish midrib. Many mutants have been confirmed in the next

generation (Pedersen JF, personal communication). Ongoing complementation study showed that in addition to many alleles of the previous known *bmr* loci, six novel mutants that could not complement the previously known loci were also identified. It is not clear how many new loci these six mutants represent. These novel mutants and new alleles of previously known loci provide new genetic resource to improve the digestibility of forage sorghum and the conversion efficiency of sorghum stover to ethanol while minimizing the effect of *bmr* mutation on biomass production and lodging.

2.3 Erect Leaf Mutants

Total biomass yield and efficient conversion of the biomass to bioenergy are two critical factors for sorghum to become a major bioenergy feedstock. Although sorghum has excellent tolerance to abiotic stresses such as drought and high temperature, and can thrive on poor soil with minimal fertilizer, sorghum biomass and grain yield are generally lower than maize across a range of environmental conditions (Mason et al. 2008). Moreover, the increase in potential yield of sorghum hybrids released in the several decades since the Green Revolution is only one-third of that of maize hybrids released in the same period of time (Dhugga 2007). Regardless of the pace of the increase, the improvement in genetic yield potential in both maize and sorghum is strongly correlated with increases in the number of ears (maize) or panicles (sorghum) per unit area. Over this period of time, the density of maize hybrids increased by an average of ~1,000 plants·per hectare per year, corresponding to ~1% annual increase in grain yield (Dhugga 2007). In the 36 maize hybrids released from 1936 to 1991, leaf angle score of new hybrids displayed an improvement of 122% over the old ones, the greatest change among all ten plant traits examined (Duvick and Cassman 1999). The modern maize hybrids have much more acute (erect) leaf angle than older hybrids, which allows the hybrids to be planted at higher density to capture more solar radiation per unit land area (Duvick and Cassman 1999). Erect leaf mutants in rice have also been shown to have increased biomass and grain yield (Sakamoto et al. 2006).

Compared with modern maize hybrids, sorghum exhibits an open canopy with wide leaf angles that almost parallel the ground. A sorghum mutant with erect leaf angle has been reported previously (Singh and Drolsom 1973). This mutant, associated with no leaf ligule and other undesirable traits, has not been used to improve leaf angle in sorghum breeding. Among the 3,000 M_3 plots in the field, over 50 plots segregated for leaf angles that vary from the wild-type BTx623. Eleven of these mutants were confirmed at the next generation (M_4). Several mutants have similar or slightly bigger panicles than wild type (Table 8.1). Although these *erl* mutants need to be confirmed in a segregating F_2 population and homozygous F_3 generation under different environments and plant densities, some may prove to be useful for improving sorghum biomass and grain production based on the yield improvement achieved in maize hybrids through improved leaf angle (Dhugga 2007). Other traits

Table 8.1 A list of erect leaf mutants confirmed at M4 generation

Line	Height (cm)	Second leaf angle (°)	Second leaf width (cm)	Second leaf length (cm)	Head length (cm)	Seed weight (g)
BTx623	163	45	7.5	56	32	78
M2P1374	96	82	9	65	34	7.6
M2P0514	101	80	6	56	25	10.2
MUT841	121	70	7.4	74.5	35	32.1
MUT1008	112	70	8.5	43	21	56.6
M2P0630	122	70	8.7	56	28	73.4
MUT1169	131	65	7	48	32	60.9
M2P0819	122	65	7	55	25	44
M2P0684	124	60	6	65	26	28.7
M2P0784	128	60	7	60	29	38.4
25M2-0552	113	60	7	49	23	12.3
20M2-0024	138	60	8.5	54	35	74

Leaf angle was measured at full bloom on the leaf below the flag leaf. A leaf parallel to ground has an angle of zero and perpendicular to ground has an angle of 90°

from the mutant library, such as monoculm, multiple tillers, and large panicle sizes (Fig. 8.1), may also help to improve biomass and grain production in sorghum. Furthermore, beneficial traits may be stacked to increase biomass yield and biomass conversion efficiency to develop feedstock genotypes tailored to bioenergy production (for example, crossing *bmr* mutants with *mtl* mutants and/or *erl* mutants to develop double or triple mutant plants).

2.4 A Sorghum Mutant Library as a Source to Identify Beneficial Traits for Other Saccharinae Species

Phenotype diversity and frequency after mutagenesis depend on the dosage of mutagen used, polyploidy level, and gene redundancy. As a minimally redundant genome that last experienced genome duplication ~70 million years ago (MYA) (Paterson et al. 2004), sorghum is very sensitive to mutagenesis treatments and displays a wide spectrum and high frequency of mutant phenotypes even after treatment with low dosage of mutagens (Peters et al. 2009; Sree Ramulu 1970a, b; Xin et al. 2008). In hexaploid and tetraploid wheat, less than 0.5% of mutagenized lines display visible mutant phenotypes even after heavy mutagenesis (0.6–1.2% EMS for 18 h) that resulted in a mutation rate of one base substitution per 24–80 kb (Slade et al. 2005). Rice, a diploid plant, has experienced an evolutionary history very similar to that of sorghum, with no genome duplication in about 70 million years (Paterson et al. 2004; Yu et al. 2005). However, rice often requires high concentrations of mutagens to create mutant populations that have useful mutation frequencies (Till et al. 2007; Wu et al. 2005). After treatment with 0.8% EMS, about 3% of the M_3 lines in rice segregated for albinism (Wu et al. 2005), while 17% of M_3 lines in sorghum treated

with 0.25% EMS segregated from albino seedlings (Xin et al. 2008). Other common mutant phenotypes occur at much lower frequency in mutagenized rice than in sorghum. The presence of a husk outside of the rice grain may account in part for the requirement of a high concentration of EMS but gamma radiation, which can penetrate husks with little impedance, results in only slightly higher albinism rates (Wu et al. 2005). To overcome the effect of husks, a mutagenesis protocol for treating individual zygotes in developing panicles has been developed (Suzuki et al. 2007). Mutation rates in populations generated with this new protocol appear to be higher at DNA level than those in previous populations generated by treating mature seeds (Suzuki et al. 2007; Till et al. 2007). The frequency of albinism or other visible mutant phenotypes was not reported for this mutagenized population. Maize, although a "diploid" plant based on chromosome pairing, is believed to have experienced a whole genome duplication just 11 MYA and thus has more gene redundancy than sorghum. TILLING populations have been developed in maize with adequate mutation frequency at DNA level (Till et al. 2004; Weil and Monde 2007); however, the frequency and diversity of visible mutant phenotypes have not been reported.

Many close relatives of sorghum are polyploids. Sugarcane (*Saccharum*) is an autopolyploid with variable number of chromosomes from $2n = 80$ to 140 (Dillon et al. 2007; Irvine 1999). *Miscanthus* × *giganteus* is a sterile allotriploid that is naturally crossed from diploid *Miscanthus sinensis* ($2n = 38$) and tetraploid *Miscanthus sacchariflorus* ($4n = 76$) (Rayburn et al. 2009). Due to the variable chromosome number and ploidy levels, no natural or induced mutant has been reported. It is likely to be challenging to develop mutant populations for these species. A sorghum mutant library with well-annotated phenotypes could identify useful traits for improving close relatives such as Saccharum and Miscanthus for biomass production and bioenergy conversion efficiency. For example, *bmr* mutations have shown to increase the digestibility and ethanol conversion efficiency of maize and sorghum stover (Sattler et al. 2010; Vermerris et al. 2007). Two genes in the lignin synthesis pathway, that lead to the *bmr* phenotype when mutated, CAD and COMT, have been cloned. RNAi techniques might be applied to other grasses such as *Miscanthus* × *giganteus* to down-regulate these two genes to improve the efficiency of converting biomass to ethanol. Due to recent genome duplications, the annotated sorghum mutant library may be also useful for functional studies of maize genes.

2.5 TILLING

Heritable mutation induced by physical radiation, transposon insertion, or chemical mutagenesis with an ideal phenotype (such as early flowering) is occasionally but infrequently used directly for agricultural production, sometimes used as a source of variation in the breeding process, and very frequently used for biological studies. Large chromosomal rearrangements (such as insertions, deletions, inversions, or translocations) usually induced by radiation can be traced by cytogenetics. However, detection of small insertions or deletions (indels) and point mutations is far beyond

the resolution of cytogenetic analysis. Targeting induced local lesions in genomes (TILLING) (McCallum et al. 2000a, b) is a new technique that can be efficiently used for detecting small indels and point mutations. TILLING is particularly suitable for genome-wide analysis with a large mutant population induced by chemical mutagenesis and can be used in a high-throughput format to make links between characterized mutant genotypes and the resulting phenotypes. TILLING mainly includes the following steps: development of a mutant population with a reasonable number of individuals by mutagenesis; collection of leaf tissue from M_2 plants and extraction of DNA; pool of DNA samples from mutants by one or multiple dimensions; design of PCR primers covering regions of interest from sequence database; PCR amplification of targeted regions of interest; heteroduplex formation between wild-type and mutated amplicons; detection of mutation by dHPLC or other separation systems; dissection of individuals within a pool containing a mutation by further PCR; reconfirmation of an individual with a mutation by sequencing; and association of the confirmed mutant genotype with a resulting mutant phenotype. The details of TILLING procedures can be altered or modified from lab to lab. For example, the chemicals used for mutagenesis, the number of mutants within a population for covering the full spectrum, and the number of pooling dimensions and folds should be dependent on species. The detection platform can be dHPLC (Underhill et al. 1997), agarose gel system (Raghavan et al. 2007), and denaturing polyacrylamide gel or even capillary system depending on the equipment available in the lab. A subset of 768 lines from the AIMS mutant library were selected to conduct a pilot TILLING (Xin et al. 2008). Despite the sensitivity of sorghum to EMS treatment, a mutation rate of 1/526 kb was found with four amplicons, a mutation rate that is adequate for high-throughput TILLING to deduce the function of sorghum genes.

3 Transposon Mutagenesis

3.1 Sorghum Candystripe1 Transposon

In sorghum, a variegated line was originally collected by Dr. O. Webster from Gedaref, Sudan. Genetic analysis and mutable behavior of this allele indicated the presence of a transposon (Zanta et al. 1994). Further molecular characterization led to the isolation of the transposable element associated with the sorghum candystripe (variegated) phenotype (Chopra et al. 1999). The host sequence of the transposon was the *yellow seed 1* (*y1*) gene which encodes an MYB domain protein that is closely related to P1 and homologous MYB proteins in maize and teosinte (Chopra et al. 2002; Jiang et al. 2004; Zhang et al. 2000). The variegated allele was designated as *y1-candystripe (y1-cs).* Based on the high frequency of somatic and germinal reversions of *y1-cs* to functional *y1*, it was obvious that the phenotype of the *y1-cs* allele results from the presence of a transposable element in the *y1* gene (Hu et al. 1991; Zanta et al. 1994). Moreover, the *y1-cs* allele bears a marked

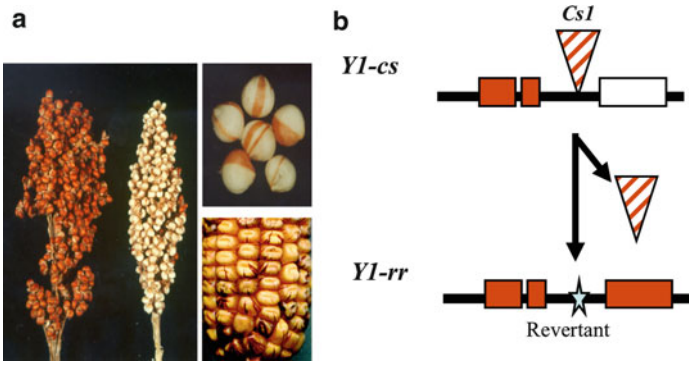

Fig. 8.2 *Panel A* depicts the Sorghum candystripe phenotype and its similarity to the maize variegated phenotype conditioned by *P1-vv*. *Left* of *panel A* shows a sorghum head with variegated and *red kernels* next to a *full red head* (revertant). *Panel B* depicts the position of the *Cs1* transposon in the *y1* sequence. Loss of the *Cs1* results in generation of a functional *y1* allele. *Star* indicates the 2 bp footprint left upon excision of the *Cs1* element

resemblance to the maize *p1-vv* allele. In maize, the *p1* gene regulates the production of the phlobaphenes in kernel pericarp and other plant tissues (Chopra et al. 1996; Lechelt et al. 1989). The *y1-cs* allele specifies variegated seed pericarp pigmentation (Fig. 8.2). By virtue of its phenotypic effect on the expression of the pericarp pigmentation, the transposon was named as *Candystripe1* (*Cs1*) (Chopra et al. 1999). It is a member of the CACTA family of plant transposable elements, which has been suggested to be an important vector for reshuffling sorghum exons and genes, similar to PACKMULEs in rice and Helitrons in maize (Paterson et al. 2009b).

Of several members of this family, the *En/Spm* element of maize has been best understood at the genetic and molecular levels. It was originally identified as Enhancer (*En*) (Peterson 1960) and was shown to be homologous to the Suppressor–Mutator system (*Spm*) (McClintock 1954) by genetic and molecular tests (Pereira et al. 1986; Peterson 1965). Besides maize, CACTA elements have been characterized from several other plant species.

3.2 Structural Features of Candystripe1 Transposon

Cs1 has features similar to those of other members of the CACTA family: a short terminal inverted repeat (TIR) sequence 5′-CACTATGTGAAAAAAGCTTA-3′, and these termini are flanked by 3-bp target site duplication. Subterminal regions, 250 bp interior to the TIR, contain multiple copies of direct and indirect repeats. A 12 bp partially conserved sequence motif [5′-TTATTACAGACG-3′] is repeated eight and six times, respectively, in the 5′- and 3′- subterminal regions. Sequences similar to the subterminal repeat motif are also present at seven sites in the central region of the *Cs1* transposon. Interspersed within the *Cs1* element are other tandem repeats as well as several copies of High Copy Short interspersed Repeats (HCSRs).

Several of these repeat sequences have high similarity (up to 95%) to an *Sb1* Tourist element of sorghum as well as to other Miniature Inverted Repeat Transposable Elements (MITEs) (Wessler et al. 1995). Transposable elements of the CACTA family are relatively large (*En/Spm*: 8.2 kb, *Tam1*: 15.1 kb) (Gierl 1996); the 23,018 bp *Cs1* element is the largest known member of this family. The *Cs1* copy present at the *y1* locus is an autonomous element (Zanta et al. 1994). Recent sequence analysis of sorghum genome has shown the presence of deletion derivatives of CACTA elements that carry pieces of other genes (Paterson et al. 2009a).

3.3 Transposition and Reinsertion of Cs1

Cs1-Y1 offers a unique genetic system to study biology and function of *Cs1* transposons in sorghum. The *y1-cs* allele can revert to a functional state (*Y1-rr*; red pericarp and red glumes; see Fig. 8.2) in both somatic and germinal tissues, resulting in the appearance of frequent red sectors and fully red seed heads, respectively. Germinal reversions lead to a heritable and functional *y1* gene, while somatic reversions are not heritable and their frequencies vary within the progeny and depend on the transposition rates of the *Cs1* element in a tissue and genetic background (Chopra et al. 2002; Zanta et al. 1994). Full red revertants (plants producing red panicles) appear at a frequency ranging between 12 and 20% in a growout of the homozygous *y1-cs* line or crosses involving *y1-cs* allele with sorghum inbreds having different backgrounds (Carvalho et al. 2005; Chopra et al. 1999; Zanta et al. 1994). Reinsertion of the excised *Cs1* elsewhere in the genome was identified in full red head plants (*Y1/Y1* or *Y1/y1-cs*) derived from excision of the *Cs1* element from the *y1-cs*. Furthermore, PCR amplification of DNA from several independent red revertant plants using flanking primers showed that the *Cs1* element excised from the *y1* gene and left a 2 bp footprint (Chopra et al. 1999).

3.4 Cs1 Can Generate Large Deletions in the Flanking Sequences

When the *Cs1* element excises from the host sequence, the majority of excision events leave a 2-bp footprint. For example, *Cs1* is inserted into intron II of the *y1* gene, and 2-bp footprints do not affect Y1-reading frame or intron splicing and thus lead to the normal function of the gene. These excisions do produce germinal red revertants with a frequency of about 20% (Carvalho et al. 2005). However, screens for germinal excision events resulted in identification of loss of *y1* function alleles (Ibraheem et al. 2010). Through DNA gel blot analysis, these excision events were found to contain partial deletions in the *y1* gene and of the *Cs1*. The mechanism of deletions in the *Cs1* and the flanking *y1* DNA sequences is not yet clear. In addition to finding deletion derivatives, collection of several *y1-cs* alleles with different degrees of variegation demonstrates alleles with differential excision activities

during plant growth and development. These alleles can be attributed to somatic or germinal excision of the *Cs1* element from the *y1* locus. Characterization of these alleles may allow answering the question of excision mechanism(s) of the *Cs1* transposon as has been demonstrated in the case of maize *Ac* at the *p1* locus (Peterson 1990). Genetic and molecular analysis of these alleles will establish multiple copies of the element (because of linked or short-range transposition) that may be producing a negative dosage effect which in turn may be responsible for silencing leading to variable degrees of transposition (Dooner and Belachew 1991).

3.5 Genome Mutagenesis Utilizing Cs1 in Sorghum

Cs1-homologous sequences are present at low copy number in sorghum (Chopra et al. 1999). The low copy number combined with high transposition frequency of *Cs1* implies that this transposon could prove to be an efficient gene isolation tool in sorghum. Additionally, there are at least 12 copies of the defective *Cs1* elements in sorghum. As opposed to the full-length sequence of the autonomous *Cs1* elements, the defective elements range in size from 400 bp to 4.0 kbp. Sequence analysis of these deletion derivatives (*dCs2, dCs3,* and *dCs4)* showed that 200 bp on the 5′ end and up to 150 bp at the 3′ end are conserved, while each copy carries varying lengths of internal deletions (Carvalho et al. 2005). Another interesting feature of *Cs1* is that its transposition appears to be sensitive to environmental conditions. Transposition of the *Antirrhinum Tam3* element is sensitive to temperature, and this provides a means to control the frequency of transposition (Harrison and Fincham 1964). Further, effect and analysis of the flanking genomic sequences to investigate the genomic context of *Cs1* elements is underway.

3.6 Characterization of Selected Cs1 Induced Mutations

A sorghum line (CS8110419) with moderate activity of *Cs1* was crossed with an agronomically well-adapted line Tx2737 and this cross was characterized further. In the F_1 generation, plants were selfed and their DNA was used to test for transposon excision and insertion activities. All randomly selected F_1 plants showed polymorphic patterns for transposon insertions that were different from the either of the two parental lines. All F_1 plants were maintained by selfing until the F_7 generation (Carvalho et al. 2005). In each generation, red revertant as well as candystripe plants were saved. Red revertant plants were further used for molecular and genetic analysis to molecular and genetic analysis to follow excision and reinsertion of the *Cs1* in the genome. Simultaneously, any mutant phenotypes were also saved and characterized in further generations. Segregating families showing mutant/wild-type phenotypes were further selected for gene identification and isolation work using *Cs1* as a tag (Fig. 8.3). Isolation of gene(s) tagged with a *Cs1* element will provide the

Fig. 8.3 *Candystripe* tagged putative mutants. (**a**) Zebra crossbands; (**b**) oldgold; (**c**) bloomless—mutant indicated with a *red arrow* and compared with the w-t plant on the *left*; (**d**) wilty; (**e**) yellow green; (**f**) lession mimic; (**g**) brown midrib; (**h**) striate leaves; (**i**) dwarf; (**j**) iojap striping; (**k**) virescent; (**l**) third leaf yellow; (**m**) premature senescence

ultimate proof of its ability as a transposon tag in sorghum. Sequences isolated from these mutants will be mapped onto the sorghum genetic map to enrich the map with respect to phenotypic markers. Additionally, genes identified through this approach will be of interest in sorghum improvement programs utilizing genetic information for lignin biosynthesis (*bmr* mutations), drought tolerance (bloomless mutations), and disease resistance (wilty and lesion mimics). Results based on the analysis of these putative mutants have further strengthened the use of the *Cs1* element in generating and selecting non-targeted insertions. Developing a two-element system in sorghum also seems feasible and efforts have been focused on this aspect. Two-element systems have proven successful both as endogenous (*Ac/Ds* or *En/dSpm* in

maize) and heterologous systems (*En/dSpm* in Arabidopsis). To find the reporter *dCs* element that functions in the presence of the autonomous *Cs1*, we have recently recovered an anthocyanin phenotype in the endosperm.

4 Perspective

The sorghum genome sequence together with its rich genetic resources (see Kresovich et al. Chap. 2, this volume) and cost-effective sequencing technology makes us rethink how to use these resources and technology for research. TILLING and EcoTILLING can mine induced mutations and natural sequence variation, respectively, within a species (Barkley and Wang 2008; Barkley et al. 2008; Comai et al. 2004). As sequence technology becomes increasingly cost-effective, it may become reasonable to sequence hundreds or more sorghum germplasm accessions as references for mining natural sequence variation. These unmutagenized and sequenced germplasm accessions can also be used as references for TILLING. An EMS-induced sorghum mutant population has been generated and used in the sorghum community (Xin et al. 2008). To broaden the mutational spectrum available, more sorghum mutagenized populations need to be generated by using different kinds of mutagens including irradiation. In the near future, TILLING by sequencing pooled mutant lines, known as SequeTILLING, will become a reality as new high-throughput and cost-effective sequencing technologies are developed (Weil 2009). A high-quality mutant library with well-annotated phenotypes and efficient pooling strategies will make SequeTILLING a very powerful approach for mutation identification.

References

Barkley N, Wang M (2008) Application of TILLING and EcoTILLING as reverse genetic approaches to elucidate the function of genes in plants and animals. Curr Genomics 9:212–226

Barkley NA, Wang ML, Gillaspie AG, Dean RE, Pederson GA, Jenkins TM (2008) Discovering and verifying DNA polymorphisms in a mung bean [*V. radiata* (L.) R. Wilczek] collection by EcoTILLING and sequencing. BMC Res Notes 1:28

Bout S, Vermerris W (2003) A candidate-gene approach to clone the sorghum Brown midrib gene encoding caffeic acid O-methyltransferase. Mol Genet Genomics 269:205–214

Carvalho C, Boddu J, Zehr U, Axtell J, Pedersen JF, Chopra S (2005) Genetic and molecular characterization of Candystripe1 transposition events in sorghum. Genetica 124:201–212

Chopra S, Athma P, Peterson T (1996) Alleles of the maize P gene with distinct tissue specificities encode Myb-homologous proteins with C-terminal replacements. Plant Cell 8:1149–1158

Chopra S, Brendel V, Zhang J, Axtell JD, Peterson T (1999) Molecular characterization of a mutable pigmentation phenotype and isolation of the first active transposable element from *Sorghum bicolor*. Proc Natl Acad Sci USA 96:15330–15335

Chopra S, Gevens A, Svabek C, Wood KV, Peterson T, Nicholson RL (2002) Excision of the *Candystripe1* transposon from a hyper-mutable *Y1-cs* allele shows that the sorghum *Y1* gene

controls the biosynthesis of both 3-deoxyanthocyanidin phytoalexins and phlobaphene pigments. Physiol Mol Plant Pathol 60:321–330
Comai L, Young K, Till BJ, Reynolds SH, Greene EA, Codomo CA, Enns LC, Johnson JE, Burtner C, Odden AR, Henikoff S (2004) Efficient discovery of DNA polymorphisms in natural populations by EcoTILLING. Plant J 37:778–786
Dhugga KS (2007) Maize biomass yield and composition for biofuels. Crop Sci 47:2211–2227
Dillon SL, Shapter FM, Henry RJ, Cordeiro G, Izquierdo L, Lee LS (2007) Domestication to crop improvement: genetic resources for *Sorghum* and *Saccharum* (Andropogoneae). Ann Bot 100:975–989
Dooner H, Belachew A (1991) Chromosome breakage by pairs of closely linked transposable elements of the Ac-Ds family in maize. Genetics 129:855
Duvick DN, Cassman KG (1999) Post-green revolution trends in yield potential of temperate maize in the North-Central United States. Crop Sci 39:1622–1630
Ejeta G, Axtell J (1985) Mutant gene in sorghum causing leaf "reddening" and increased protein concentration in the grain. J Hered 76:301–302
Ellstrand NC, Foster KW (1983) Impact of population structure on the apparent outcrossing rate of grain sorghum (*Sorghum bicolor*). Theor Appl Genet 66:323–327
Gaul H (1964) Mutations in plant breeding. Rad Bot 4:155–232
Gierl A (1996) The En/Spm transposable element of maize. Curr Top Microbiol Immunol 204:145–160
Greene EA, Codomo CA, Taylor NE, Henikoff JG, Till BJ, Reynolds SH, Enns LC, Burtner C, Johnson JE, Odden AR, Comai L, Henikoff S (2003) Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in Arabidopsis. Genetics 164:731–740
Harrison B, Fincham J (1964) Instability at the Pal locus in *Antirrhinum majus*. Heredity 19:237–258
Henikoff S, Comai L (2003) Single-nucleotide mutations for plant functional genomics. Annu Rev Plant Biol 54:375–401
Hu G, Kofoid K, Liang G (1991) An unstable mutation for pigmentation in kernels of 'Calico' sorghum. Hereditas 115:163–167
Ibraheem F, Gaffoor I, Chopra S (2010) Induction of 3-deoxyanthocyanidin phytoalexin dependent resistance to anthracnose leaf blight requires a functional *yellow seed1* in *Sorghum bicolor*. Genetics 184:915–926
Irvine JE (1999) Saccharum species as horticultural classes. Theor Appl Genet 98:186–194
Jenks MA, Joly RJ, Peters PJ, Rich PJ, Axtell JD, Ashworth EN (1994) Chemically induced cuticle mutation affecting epidermal conductance to water vapor and disease susceptibility in *Sorghum bicolor* (L.) Moench. Plant Physiol 105:1239–1245
Jiang C, Gu J, Chopra S, Gu X, Peterson T (2004) Ordered origin of the typical two- and three-repeat Myb genes. Gene 326:13–22
Lechelt C, Peterson T, Laird A, Chen J, Dellaporta SL, Dennis E, Peacock WJ, Starlinger P (1989) Isolation and molecular analysis of the maize P locus. Mol Gen Genet 219:225–234
Mason SC, Kathol D, Eskridge KM, Galusha TD (2008) Yield increase has been more rapid for maize than for grain Sorghum. Crop Sci 48:1560–1568
McCallum CM, Comai L, Greene EA, Henikoff S (2000a) Targeted screening for induced mutations. Nat Biotechnol 18:455–457
McCallum CM, Comai L, Greene EA, Henikoff S (2000b) Targeting induced local lesions IN genomes (TILLING) for plant functional genomics. Plant Physiol 123:439–442
McClintock B (1954) Mutations in maize and chromosomal aberrations in Neurospora. Carnegie Inst Wash Year Book 53:254–260
Menz MA, Klein RR, Mullet JE, Obert JA, Unruh NC, Klein PE (2002) A high-density genetic map of *Sorghum bicolor* (L.) Moench based on 2926 AFLP, RFLP and SSR markers. Plant Mol Biol 48:483–499
Oria MP, Hamaker BR, Axtell JD, Huang CP (2000) A highly digestible sorghum mutant cultivar exhibits a unique folded structure of endosperm protein bodies. Proc Natl Acad Sci USA 97:5065–5070

Paterson AH (2008) Genomics of sorghum. Int J Plant Genomics 2008:362451
Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc Natl Acad Sci USA 101: 9903–9908
Paterson A, Bowers J, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A (2009a) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556
Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboobur R, Ware D, Westhoff P, Mayer KF, Messing J, Rokhsar DS (2009b) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556
Pereira A, Cuypers H, Gierl A, Schwarz-Sommer Z, Saedler H (1986) Molecular analysis of the En/Spm transposable element system of *Zea mays*. EMBO J 5:835
Peters PJ, Jenks MA, Rich PJ, Axtell JD, Ejeta G (2009) Mutagenesis, selection, and allelic analysis of epicuticular wax mutants in Sorghum. Crop Sci 49:1250–1258
Peterson PA (1960) The pale green mutable system in maize. Genetics 45:115–133
Peterson P (1965) A relationship between the Spm and En control systems in maize. Am Nat 99:391–398
Peterson T (1990) Intragenic transposition of Ac generates a new allele of the maize P gene. Genetics 126:469
Porter KS, Anxtell JD, Lechtenberg VL, Colenbrander VF (1978) Phenotype, fiber composition, and in vitro dry matter disappearance of chemically induced brown midrib (bmr) mutants of sorghum. Crop Sci 18:205–208
Quinby JR (1975) The genetics of Sorghum improvement. J Hered 66:56–62
Quinby JR, Karper RE (1942) Inheritance of mature plant characters in Sorghum: induced by radiation. J Hered 33:323–327
Raghavan C, Naredo M, Wang H, Atienza G, Liu B, Qiu F, McNally K, Leung H (2007) Rapid method for detecting SNPs on agarose gels and its application in candidate gene mapping. Mol Breed 19:87–101
Rayburn A, Crawford J, Rayburn C, Juvik J (2009) Genome size of three *Miscanthus* species. Plant Mol Biol Rep 27:184–188
Saballos A, Ejeta G, Sanchez E, Kang C, Vermerris W (2009) A genomewide analysis of the cinnamyl alcohol dehydrogenase family in sorghum [*Sorghum bicolor* (L.) Moench] identifies SbCAD2 as the brown midrib6 gene. Genetics 181:783–795
Sakamoto T, Morinaka Y, Ohnishi T, Sunohara H, Fujioka S, Ueguchi-Tanaka M, Mizutani M, Sakata K, Takatsuto S, Yoshida S, Tanaka H, Kitano H, Matsuoka M (2006) Erect leaves caused by brassinosteroid deficiency increase biomass production and grain yield in rice. Nat Biotechnol 24:105–109
Sattler SE, Saathoff AJ, Haas EJ, Palmer NA, Funnell-Harris DL, Sarath G, Pedersen JF (2009) A nonsense mutation in a cinnamyl alcohol dehydrogenase gene is responsible for the Sorghum brown midrib6 phenotype. Plant Physiol 150:584–595
Sattler SE, Funnell-Harris DL, Pedersen JF (2010) Brown midrib mutations and their importance to the utilization of maize, sorghum, and pearl millet lignocellulosic tissues. Plant Sci 178:229–238
Singh R, Axtell JD (1973) High lysine mutant gene (hl that improves protein quality and biological value of grain Sorghum. Crop Sci 13:535–539
Singh SP, Drolsom PN (1973) Induced recessive mutations affecting leaf angle in *Sorghum bicolor*. J Hered 64:65–68
Singh SP, Drolsom PN (1974) Induced early-maturing mutation in Sorghum. Crop Sci 14:377–380
Slade AJ, Fuerstenberg SI, Loeffler D, Steine MN, Facciotti D (2005) A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. Nat Biotechnol 23:75–81

Sree Ramulu K (1970a) Induced systematic mutations in Sorghum. Mutat Res Fund Mol Mech Mutagen 10:77–80

Sree Ramulu K (1970b) Sensitivity and induction of mutations in sorghum. Mutat Res Fund Mol Mech Mutagen 10:197–206

Sree Ramulu K, Sree Rangasamy SR (1972) An estimation of the number of initials in grain Sorghum using mutagenic treatments. Rad Bot 12:37–43

Suzuki T, Eiguchi M, Kumamaru T, Satoh H, Matsusaka H, Moriguchi K, Nagato Y, Kurata N (2007) MNU-induced mutant pools and high performance TILLING enable finding of any gene mutation in rice. Mol Genet Genomics 279:213–223

Till BJ, Reynolds SH, Weil C, Springer N, Burtner C, Young K, Bowers E, Codomo CA, Enns LC, Odden AR, Greene EA, Comai L, Henikoff S (2004) Discovery of induced point mutations in maize genes by TILLING. BMC Plant Biol 4:12

Till BJ, Cooper J, Tai TH, Colowit P, Greene EA, Henikoff S, Comai L (2007) Discovery of chemically induced mutations in rice by TILLING. BMC Plant Biol 7:1

Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. Genome Res 7:996–1005

Vermerris W, Saballos A, Ejeta G, Mosier NS, Ladisch MR, Carpita NC (2007) Molecular breeding to enhance ethanol production from corn and sorghum stover. Crop Sci 47:S-142–S-153

Weil CF (2009) TILLING in grass species. Plant Physiol (Lancaster, PA) 149:158–164

Weil CF, Monde R-A (2007) Getting the point-mutations in maize. Crop Sci 47:S-60–S-67

Wessler S, Bureau T, White S (1995) LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. Curr Opin Genet Dev 5:814–821

Wu JL, Wu C, Lei C, Baraoidan M, Bordeos A, Madamba MR, Ramos-Pamplona M, Mauleon R, Portugal A, Ulat VJ, Bruskiewich R, Wang G, Leach J, Khush G, Leung H (2005) Chemical- and irradiation-induced mutants of indica rice IR64 for forward and reverse genetics. Plant Mol Biol 59:85–97

Xin Z, Wang ML, Barkley NA, Burow G, Franks C, Pederson G, Burke J (2008) Applying genotyping (TILLING) and phenotyping analyses to elucidate gene function in a chemically induced sorghum mutant population. BMC Plant Biol 8:103

Yu J, Wang J, Lin W, Li S, Li H, Zhou J, Ni P, Dong W, Hu S, Zeng C, Zhang J, Zhang Y, Li R, Xu Z, Li X, Zheng H, Cong L, Lin L, Yin J, Geng J, Li G, Shi J, Liu J, Lv H, Li J, Deng Y, Ran L, Shi X, Wang X, Wu Q, Li C, Ren X, Li D, Liu D, Zhang X, Ji Z, Zhao W, Sun Y, Zhang Z, Bao J, Han Y, Dong L, Ji J, Chen P, Wu S, Xiao Y, Bu D, Tan J, Yang L, Ye C, Xu J, Zhou Y, Yu Y, Zhang B, Zhuang S, Wei H, Liu B, Lei M, Yu H, Li Y, Xu H, Wei S, He X, Fang L, Huang X, Su Z, Tong W, Tong Z, Ye J, Wang L, Lei T, Chen C, Chen H, Huang H, Zhang F, Li N, Zhao C, Huang Y, Li L, Xi Y, Qi Q, Li W, Hu W, Tian X, Jiao Y, Liang X, Jin J, Gao L, Zheng W, Hao B, Liu S, Wang W, Yuan L, Cao M, McDermott J, Samudrala R, Wong GK, Yang H (2005) The genomes of *Oryza sativa*: a history of duplications. PLoS Biol 3:e38

Zanta C, Yang X, Axtell J, Bennetzen J (1994) The candystripe locus, y-cs, determines mutable pigmentation of the Sorghum leaf, flower, and pericarp. J Hered 85:23

Zhang P, Chopra S, Peterson T (2000) A segmental gene duplication generated differentially expressed myb-homologous genes in maize. Plant Cell 12:2311–2322

Chapter 9
Association Genetics Strategies and Resources

Jianming Yu, Martha T. Hamblin, and Mitchell R. Tuinstra

Abstract Genomic technologies are making it possible to strategically exploit genetic diversity in crops to map complex agronomic and physiological traits and improve these traits for grain and biomass production. Sorghum is well positioned to benefit from these association genetics strategies, and essential components of association mapping have been established. Research in sorghum association mapping contributes to a better understanding of genetics of complex traits and improved breeding methods to exploit genetic diversity.

Keywords Association mapping • Candidate gene • Complex trait dissection • Genetic diversity • Genome-wide association study • Linkage disequilibrium • Nested association mapping • Single nucleotide polymorphism • Sorghum diversity panel

1 Association Genetics

Many genetic mapping studies in plants have been conducted with recombinant inbred line (RIL) populations from a biparental cross because it is easy to maintain these populations for replicated trials (Bernardo 2008; Holland 2007). In contrast, association genetics has been implemented extensively in human genetics studies,

J. Yu (✉)
Department of Agronomy, Kansas State University,
3004 Throckmorton Plant Science Center, Manhattan, KS 66506-5501, USA
e-mail: jyu@ksu.edu

M.T. Hamblin
Institute for Genomic Diversity, Cornell University, Ithaca, NY, USA
e-mail: mth3@cornell.edu

M.R. Tuinstra
Department of Agronomy, Purdue University, West Lafayette, IN, USA
e-mail: mtuinstr@purdue.edu

A.H. Paterson (ed.), *Genomics of the Saccharinae*, Plant Genetics and Genomics: Crops and Models 11, DOI 10.1007/978-1-4419-5947-8_9,

partly because of the early adoption of large-scale genotyping strategies and the necessity of exploiting population-based samples for studying complex human diseases. However, widespread use of single nucleotide polymorphism (SNP) markers and the reduced cost of sequencing and genotyping have led researchers working with different plant species to adopt association mapping and the underlying linkage disequilibrium (LD) approach (Zhu et al. 2008).

Here we briefly introduce the concept of linkage analysis and association mapping. Readers should refer to other detailed reviews for a full explanation (Flint-Garcia et al. 2003; Nordborg and Tavare 2002; Risch and Merikangas 1996; Zhu et al. 2008). In essence, both linkage analysis and association mapping strategies are designed to identify marker–trait association signals that result from co-inheritance of functional polymorphisms and neighboring DNA variants (markers). In linkage analysis in plants, the signals are typically generated by co-inheritance within a segregating population. This segregating population starts with the cross of two homogenous inbred parents and contains one or more generations of recombination. Association mapping is aimed at detection of marker–trait association signals within a broad collection of accessions—natural populations, landraces, breeding lines, or a combination of these. The reasoning behind this approach is that historical recombinations and genetic diversity captured in this collection would allow fine map resolution because any markers that are not tightly linked to true functional polymorphisms would not generate any strong signals (Risch and Merikangas 1996; Zhu et al. 2008).

2 Association Genetics in Plants

Several recent review and perspective papers have documented the current status of association mapping and pointed out challenges that need to be addressed in plants (Myles et al. 2009; Nordborg and Weigel 2008; Zhu et al. 2008). Nevertheless, because genome sequence projects have been completed for the model plant species Arabidopsis (The Arabidopsis Genome Initiative 2000) and several major crops, including rice (Goff et al. 2002; Yu et al. 2002), sorghum (Paterson et al. 2009), maize (Schnable et al. 2009), and soybean (Schmutz et al. 2010), and because genomes of many other plants are being sequenced, association genetics, both as a general strategy in complex trait dissection and as a complementary approach to other existing tools, is expected to attract further attention.

A recent review documented association genetics studies in plant species such as maize, Arabidopsis, sorghum, wheat, barley, potato, rice, loblolly pine, sugarcane, eucalyptus, and perennial ryegrass (Zhu et al. 2008). Association mapping panels in different crops have been established as community resources, and findings from these studies are promising. Regardless of the degree of LD in mapping panels of different crops—for example, fast decay in diverse maize lines in thousands of base-pairs (Yu and Buckler 2006) or slow decay in wheat breeding lines in centimorgans (Sorrells and Yu 2009)—association genetics has been embraced as a powerful addition to the genetic analysis toolbox.

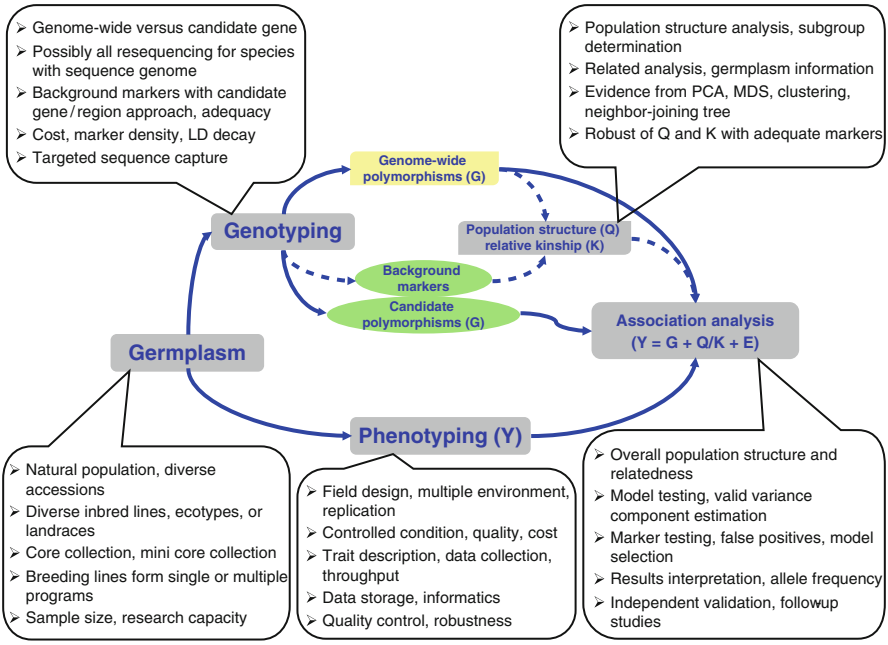

Fig. 9.1 General procedures of association mapping and highlights of different steps

2.1 Association Mapping Procedures

General components of association mapping include germplasm, genotyping, phenotyping, and analysis (Fig. 9.1) (Zhu et al. 2008). Unlike linkage mapping, association mapping usually involves assembling a collection (or population but without referring to segregating populations) of ready-to-measure accessions or lines rather than developing an F_2, BC, or RIL population. A newer strategy that combines features of association mapping with diverse lines and linkage analysis with segregating populations will be introduced in Sect. 2.4.

Genotyping in association mapping also represents a significant departure from traditional linkage analysis. The marker density requirement for a robust analysis is generally much higher for association mapping than for linkage analysis even though low-density genotyping with random markers across an association mapping panel mimics the genotyping process of traditional linkage analysis with an F_2, BC, or RIL population. For association mapping, such effort is primarily geared toward assessing population structure and genetic relatedness of a collection by examining the marker information collectively (Yu et al. 2009), not toward testing these markers individually for marker–trait association unless a very significant number of markers are used (Zhu et al. 2008). As we explain in the next two sections, markers to be tested for marker–trait association could be from candidate

genes, regions implicated in previous linkage mapping studies, or a large number of markers across the whole genome.

A well-designed association panel will have extensive phenotypic diversity. Phenotyping of such diverse materials is challenging given the broad variation in photoperiod sensitivity, flowering time, and market type and other existing differentiation within the collection (Myles et al. 2009; Zhu et al. 2008). As a result, field design, appropriate blocking, timing of record taking, data analysis, and interpretation of results all demand more effort. Data analysis for association mapping involves (1) marker data analysis such as population structure (Q), relative kinship (K), principal component analysis (P), or multidimensional scaling analysis (M); (2) trait data processing such as multiple environment data analysis; (3) model testing for appropriate models (i.e., Q, P, M, K, QK, PK, and MK); and (4) marker–trait association testing. Readers should refer to recent research and review articles for detailed information on algorithms and software packages that are commonly used in plants (Bradbury et al. 2007; Yu et al. 2006; Zhu et al. 2008).

2.2 Candidate Gene Association Mapping

In association mapping, candidate genes or regions can be targeted on the basis of metabolic and biochemical pathways, mutational studies, linkage analysis results, and genome sequence annotations from either the species of interest or relevant models (Zhu et al. 2008). This is a trait-specific, hypothesis-driven approach. As we stressed in Sect. 2.1, an adequate number of background markers need to be genotyped and analyzed for population structure and relative kinship to ensure that tests of candidate gene SNPs are valid. Recent examples of candidate gene association mapping include carotenoids in maize (Harjes et al. 2008; Yan et al. 2010) and eating and cooking properties in rice (Tian et al. 2009). In these studies, well-characterized pathways provided excellent starting points for candidate gene selection.

2.3 Genome-Wide Association Study

Association mapping can be conducted by genotyping all individuals with tens of thousands of SNPs instead of focusing on candidate genes or regions. Genome-wide association studies have been extensively conducted to dissect the genetic causes of complex human diseases for many years (Manolio et al. 2009; Wang et al. 2005). But for plants, Genome-Wide Association Study (GWAS) at a decent scale have been completed only in the model plant species Arabidopsis (Atwell et al. 2010; Zhao et al. 2007). Genome-wide association studies represent an important advance from candidate gene studies or family-based linkage studies. Large-scale GWAS typically validate findings of previously identified genes and generate new signals and hypotheses for further investigation. However, results from GWAS have also

raised some concerns about the potential limitation of association mapping, termed "missing heritability." The classic example of missing heritability is the mapping of human height (Manolio et al. 2009). Forty loci have been implicated in controlling adult height variation, but together they explain only 5% of phenotypic variation even though the estimated heritability of this trait is about 80% (Visscher 2008). Potential causes of this problem include rare allele frequency, epistasis, sample size, structure variants, and the interaction between genotype and environment. Strategies and methods in GWAS are evolving to address these concerns. It is still too early to know whether GWAS in plants will be subject to the same concerns.

2.4 Nested Association Mapping

Nested association mapping (NAM) is a special case of joint linkage and linkage disequilibrium mapping and is well suited for many plant species (Yu et al. 2008). The essence of NAM is to combine the merits of linkage analysis with designed populations and association mapping with assembled germplasm. First, a set of diverse inbred lines is selected as founders and crossed according to genetic designs (e.g., Reference Design, Design I, Design II, Diallel, Single Round Robin, or Double Round Robin). Then, RIL are developed from each cross. Genotyping the founders and RIL with a smaller set of tagging markers makes it possible to track the recombination of chromosome segments. Further, genotyping the founders with a much larger set of markers permits this marker information to be projected onto the tagged chromosome segments of the RIL. Finally, the projected marker data are combined with the phenotype data of the RIL for high-resolution mapping (Yu et al. 2008). NAM is considered a major tool for next-generation genetics (Nordborg and Weigel 2008). Multiple-family analysis of maize NAM populations provided tremendous power and precision in revealing the multigene nature of flowering time (Buckler et al. 2009; McMullen et al. 2009).

3 Resources and Examples in Sorghum Association Genetics

Over the past decade, several groups have created resources for association mapping in sorghum. These resources include carefully selected, diverse germplasm collections characterized for population structure on the basis of variation at genome-wide molecular markers as well as experimental mapping populations for joint linkage and linkage disequilibrium studies. Thanks to collaborative efforts among some of these groups, very valuable germplasm and marker resources are or will soon be publicly available for sorghum, opening the door to the integrated study of sorghum and maize and allowing incorporation of new genetic resources into sorghum breeding programs.

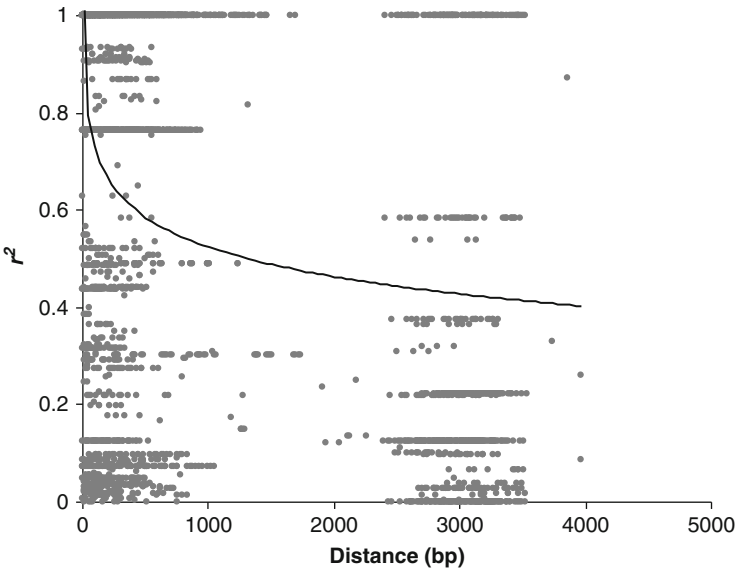

Fig. 9.2 Short-range linkage disequilibrium (r^2) as a function of distance. Data were pooled from six unlinked regions (Hamblin et al. 2005)

3.1 Linkage Disequilibrium in Sorghum

The extent of linkage disequilibrium is a key factor in the design and implementation of association genetics strategies. Given sorghum's lower estimates of sequence variation, which implies a smaller effective population size, as well as its predominately self-pollinating mating system, LD in sorghum was expected to be more extensive than in maize (Hamblin et al. 2004). This expectation was confirmed in several studies that used resequencing data from diverse sorghum lines and showed strong LD between sequence polymorphisms (i.e., SNPs) within gene-sized regions. An early study examined LD within six unlinked regions ranging in size from 40 to 100 kb and estimated the population recombination parameter, $4N_e r$, also called ρ (Hamblin et al. 2005). This parameter is useful because it summarizes LD across an entire region and has an expected relationship with the important parameter r^2: $E[r^2] = 1/(1 + 4N_e r)$. Estimates of ρ corresponded to expected r^2 values ranging from 0.14 to 0.71 for loci 10 kb apart with an average expected r^2 value of 0.25. At a distance of 1 kb, the average expected r^2 value was close to 0.8.

Although expected values provide useful information, it is critical to realize that there is tremendous variation in LD patterns across the genome. This is especially true in sorghum, a species that has recently experienced demographic events (e.g., domestication) that have dramatically perturbed its patterns of variation (Hamblin et al. 2006). Furthermore, the domestication bottleneck has had the effect of creating perfect LD (i.e., $r^2 = 1$) for sets of SNPs that have different mutational histories. For sets of SNPs that are closely linked, there has not been sufficient recombination to break up those haplotypes (Fig. 9.2).

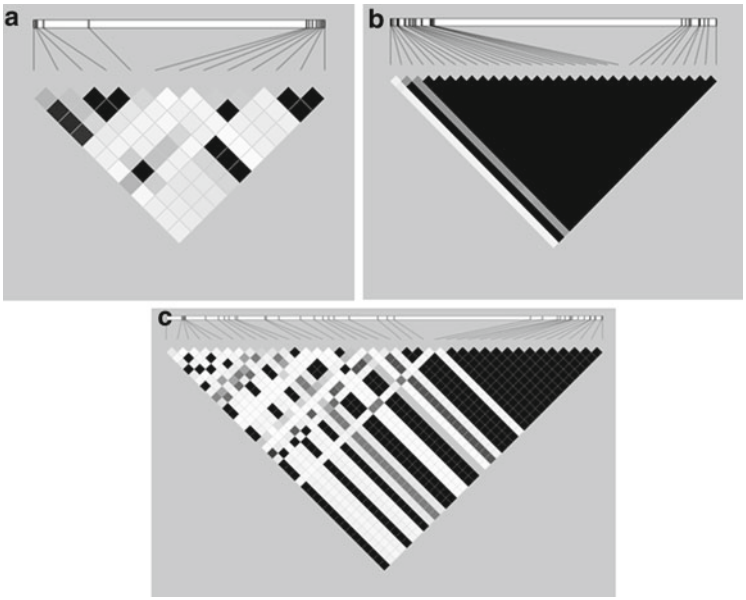

Fig. 9.3 Patterns of linkage disequilibrium vary for different genome regions. (**a**) Starch synthase III (Sb07g005400), 10 kb span; (**b**) debranching enzyme (Sb06g001540), 12 kb span; and (**c**) glucose phosphate transferase (Sb07g005200), 3 kb span. Figures were made in Haploview (Barrett et al. 2005). *Color* indicates the value of r^2 (*white* = low, *black* = high)

In another study, 15 genes in the starch metabolism pathway were sequenced in 23 lines, mostly diverse cultivars (Hamblin et al. 2007). Within 11 genes that each spanned up to 12 kb, more than 40% of SNP pairs were significantly associated at the 0.05 significance level. Haplotype structure was strong in most genes, and recombination was evident only in five genes. However, LD patterns varied widely across these regions (Fig. 9.3). In a similar study conducted to survey variation in six genes in a much larger number of lines ($N = 129–184$), little evidence of recombination was found (de Alencar Figueiredo et al. 2008).

The euchromatic regions in sorghum, which account for 97% of the genetic map, total about 252 Mb in length (Paterson et al. 2009). These LD studies suggest that a marker density for GWAS in diverse sorghum of 1 per kb in euchromatic regions, or about 250–300k SNPs, should be adequate. On the other hand, resequencing studies have revealed that there are regions in the genome where no common SNPs occur over distances of several thousand basepairs or greater. In these regions, which either have experienced selection or simply contain low variation because of genetic drift, we will have limited ability to identify genetic markers and subsequent marker–trait association signals. This is also true for centromeric regions.

3.2 Sorghum Diversity Panels

In recent years, several papers have been published that report the collection and characterization of diverse sorghum germplasm collections designed explicitly for use in association genetic studies. The Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD) assembled a core collection of 210 landraces that are representative of race, latitude of origin, response to day length, and production system and characterized them with restricted fragment length polymorphism (RFLP) probes (Deu et al. 2006). These lines represent a subset of a larger collection that was developed by the Generation Challenge Program (GCP).

A collaboration of US institutions assembled a collection of 377 accessions that represent species-wide diversity for panicle architecture and other morphological features (Casa et al. 2008). To facilitate phenotypic characterization in temperate regions, this collection was purposely assembled with accessions from sorghum conversion program and lines of historical importance in sorghum breeding. The whole collection was characterized with 47 simple sequence repeat (SSR) markers. Although the population structure is specific to the particular composition of the population, the general patterns of population structure identified in these two studies were similar. As expected, the genetic clusters correspond to the geographic and racial groupings identified in many diversity studies.

While population structure in sorghum is not strong in comparison with other self-pollinating crops like barley and rice, it is sufficient to generate modest levels of long-range LD due to admixture in panels of diverse lines, which can lead to spurious associations with phenotypes. Preliminary analysis of the US sorghum diversity panel indicated that current mixed-model methodology, accounting for both population structure and relative kinship, can adequately control for the level of population structure present in these panels (Casa et al. 2008).

A GCP-funded consortium studying genetic factors underlying drought and aluminum tolerance in sorghum is using a combined panel that includes most of the lines from the CIRAD and US panels. This consortium, led by researchers at the United States Department of Agriculture (USDA) and the Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) (Brazilian Enterprise for Agricultural Research), is also developing markers to carry out a GWAS (see Sect. 3.5). Initial genotyping with a low density of SNP markers is sufficient to find only a tiny fraction of QTL; however, LD studies with these marker sets are beginning to provide a detailed view of LD in this set of germplasm, revealing how many more markers will be necessary for whole genome coverage. This higher coverage will likely be obtained through genotyping-by-sequencing technology rather than array-based SNP genotyping technology.

The combined CIRAD-US panel of 480 lines will be made publicly available through the Germplasm Resources Information Network (GRIN), the USDA germplasm system (the US panel is already available as Sorghum Association Panel). Marker data for each line will also be made publicly available, providing a

resource for further association studies of the wide phenotypic variation captured in this collection.

A research group in Japan has also assembled a sorghum collection from 3,500 sorghum lines preserved at Genebank, National Institute of Agrobiological Science (NIAS), Japan. These lines are primarily from Asian and African sources. From an initial set of 320 lines selected on the basis of geographic distribution, 107 were chosen on the basis of diversity at 38 SSR markers (Shehzad et al. 2009b). Because this core collection is drawn from such a small germplasm collection (the US and CIRAD panels are both drawn from collections of more than 36,000 accessions) and the final size is also smaller than the US and CIRAD panels, the reduced diversity level in this NIAS panel is not unexpected. Structure analysis suggested that the NIAS lines came from three subpopulations, whereas the CIRAD and US populations appear to form nine or ten clusters. Although the NIAS population is quite small and LD is not especially extensive, QTL for 12 morphological traits were detected (Shehzad et al. 2009a).

Finally, the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) has developed a sorghum mini core collection of 242 accessions that can potentially be used for association mapping (Upadhyaya et al. 2009). This mini core collection was selected from a core collection of 2,247 accessions with phenotypic data measured for 11 qualitative traits and 10 quantitative traits. Genotyping the mini core collection and the core collection and extensively phenotyping the mini core collection would be the next steps toward using the diversity captured in these collections and inferring marker–trait associations.

3.3 Sweet Sorghum Diversity Panel

There is an emerging emphasis on using sorghum as a dedicated bioenergy crop (Carpita and McCann 2008; Rooney et al. 2007), and two sweet sorghum diversity panels have been studied (Murray et al. 2009; Wang et al. 2009). In the first study, a panel of 125 sorghum accessions was genotyped with 47 SSRs and 322 SNPs and phenotyped for brix and plant height (Murray et al. 2009). Population structure analysis indicated this panel contains three major groups of sorghum accessions: historical and modern syrup, modern sugar/energy types, and amber types. In the second study, 96 sweet sorghum accessions, an initial sample from the US historic sweet sorghum collection, were genotyped with 95 SSRs and phenotyped for flowering time, plant height, and brix (Wang et al. 2009). Although molecular marker analyses revealed weak population differentiation among these 96 accessions, the combined assessment and model testing of these three phenotypes demonstrated that this sweet sorghum panel can be classified as a type II association sample with a low level of relatedness (Zhu and Yu 2009). Genotyping experiments that include additional accessions from the US historic sweet sorghum collection are currently being conducted to expand these efforts.

3.4 Sorghum NAM Panel

On the basis of results from population structure analysis of 377 sorghum accessions using 47 SSR markers (Casa et al. 2008) and breeders' knowledge about these accessions, 10 diverse founders (SC283, SC1103, Segaolane, Macia, SC35, Ajabsido, SC971, SC265, SC1345, and P898012) were chosen from different subpopulations and crossed to the common parent, RTx430, to create the sorghum NAM population (Table 9.1). From each cross, 200 RIL were derived to form a sorghum NAM panel with 2,000 RIL (Fig. 9.4). A complementary set of RIL were planned to be derived from the crosses of the common parent Tx623 with ten different diverse founders (WL Rooney, personal communication). Each of these lines represents a subgroup identified for the sorghum diversity panel (Casa et al. 2008).

Tx430 (Miller 1984) has been widely used as a pollinator parent to produce sorghum hybrids in the USA; it is amenable to genetic engineering though both microprojectile bombardment and Agrobacterium approaches. Segeolane is a drought-tolerant kafir-type sorghum from Southern Africa (Gowda et al. 2009). Macia is a food-grade sorghum cultivar developed and selected in Tanzania for its early maturity characteristics and excellent taste attributes. Macia is high yielding and possesses preferred traits such as cooking quality and malt production characteristics for use in brewing (Bucheyeki et al. 2010). SC35 is drought resistant and has been used as a staygreen trait donor in sorghum breeding programs in the USA and Australia. Ajabsido is from Sudan and possesses excellent pre-flowering drought tolerance (Gowda et al. 2009). P898012 is well adapted to production environments in Niger and Sudan; it has both pre-flowering and post-flowering drought resistance and is amenable to transformation by both microprojectile bombardment (Casas et al. 1993) and Agrobacterium (Zhao et al. 2000). SC283 is a conspicuum-type sorghum from Tanzania and it expresses excellent tolerance to acid soils and aluminum toxicity (Bernai and Clark 1998). The rest of the NAM parents (SC1103, SC971, SC265, and SC1345) are from the sorghum conversion program but have not been well documented in literature.

3.5 SNP Genotyping Array

Using SNPs discovered in Sanger resequencing studies of more than 300 loci in samples of 16–30 sorghum accessions, Hamblin (unpublished) designed 384-SNP genotyping assays using the Illumina GoldenGate platform. These SNPs represented about 220 loci including several candidate genes each with several SNPs. More than 80% of the assays were successful. These data have been used in candidate gene-based association studies of stem sugar (Murray et al. 2009) and endosperm carotenoid content (Salas-Fernandez, unpublished) and in a study of population structure (Brown and Myles, unpublished).

The GCP consortium and the Sorghum Translational Genomics Program at Kansas State University worked collaboratively to discover additional SNPs and

Table 9.1 Diverse founders were chosen from different subpopulations and crossed to a common parent, RTx430, to create the sorghum NAM population

Accession	Subpopulation designation	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
SC283	Guinea/caudatum/bicolor (E. Africa/India)	0.835	0.003	0.003	0.081	0.003	0.004	0.013	0.004	0.007	0.011
SC1103	Sudanense/broomcorn/guinea	0.009	0.951	0.010	0.004	0.005	0.003	0.005	0.003	0.006	0.004
Segaolane	Kafir	0.017	0.004	0.896	0.003	0.002	0.009	0.034	0.014	0.016	0.004
Macia	Zerazera/caudatum	0.006	0.003	0.005	0.950	0.006	0.003	0.005	0.003	0.010	0.009
SC35	Durra	0.004	0.003	0.002	0.003	0.967	0.008	0.004	0.005	0.003	0.002
Ajabsido	Milo/feterita	0.004	0.002	0.002	0.007	0.002	0.964	0.003	0.004	0.006	0.005
SC971	Caudatum/bicolor	0.036	0.004	0.021	0.003	0.003	0.008	0.912	0.005	0.005	0.004
SC265	Guinea/caudatum (W. Africa)	0.005	0.006	0.004	0.003	0.003	0.003	0.004	0.959	0.007	0.007
SC1345	Caudatum	0.004	0.002	0.002	0.005	0.004	0.004	0.004	0.005	0.958	0.011
P898012	Caudatum/kafir	0.003	0.003	0.004	0.011	0.003	0.005	0.004	0.003	0.013	0.951
RTx430	Common parent, milo/feterita	0.006	0.022	0.008	0.410	0.003	0.49	0.007	0.031	0.006	0.017

Values are based on population structure analysis with 47 SSR markers genotyped across 377 sorghum accessions (Casa et al. 2008)

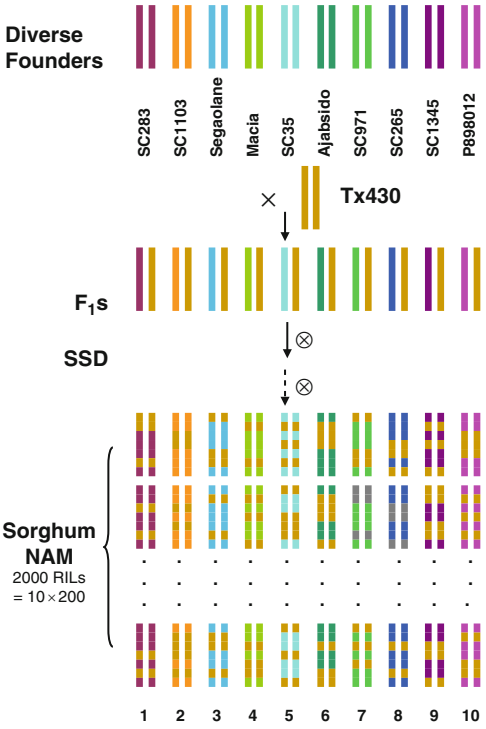

Fig. 9.4 Schematic diagram of sorghum NAM population development. The genome of each founder is *color coded* to show that the genome of RIL is a mosaic of founder genome segments. High-density genotyping of founders permits the linkage disequilibrium information captured in these diverse founders to be exploited for high-resolution mapping

develop a genotyping platform with higher density. Through this effort, Solexa sequencing of reduced representation libraries from 14 sorghum accessions, including the sorghum NAM parents, was used to discover about 34,000 high-quality, non-singleton SNPs. Discovery of SNPs in sorghum is less problematic than in maize because of the much lower level of gene duplication; most SNPs called in this analysis aligned to unique locations in the genome.

A genotyping array with 1,536 SNPs was designed to achieve maximal genome coverage (Fig. 9.5). Aside from the centromeric regions, which are very poorly represented in the SNP data, the average distance between SNPs is about 400 kb. The 480-line CIRAD-U.S. panel has been genotyped with these 1,536 markers. Much of the sorghum genome will need a much higher density of markers if we are to detect genes of modest effect underlying complex traits. This will be likely accomplished by genotyping-by-sequencing, which is quickly becoming much more cost-effective than SNP genotyping platforms such as the GoldenGate assay.

3.6 Examples of Sorghum Association Mapping

Using the US sorghum diversity panel, Brown et al. (Brown et al. 2008) examined dwarfing gene Dw3 for its association with reduced lower internode length and

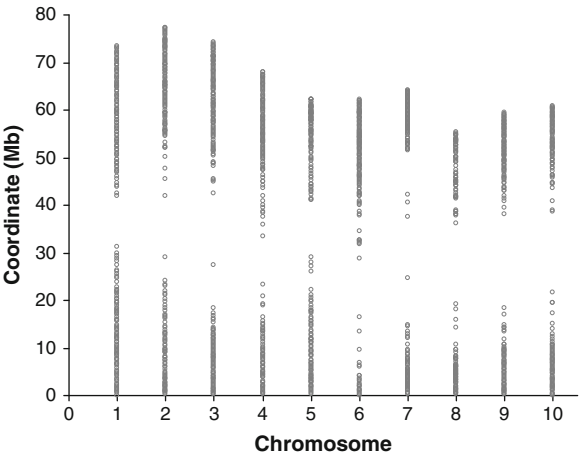

Fig. 9.5 Distribution of 1,536 SNPs across ten sorghum chromosomes

elongated apex. Fine mapping of an additional dwarfing QTL, which showed epistatic effect with Dw3, successfully narrowed the region to approximately 100 kb. In another recent study, several genomic regions associated with brix and plant height were identified (Murray et al. 2009). However, the marker density in that study (47 SSRs and 322 SNPs) was still low. Further genotyping and analysis would provide additional evidence for the detected signals.

4 Opportunities and Challenges

Essential components for carrying out large-scale association mapping studies in sorghum are in place. First, several diversity panels have been established and characterized with low-density background markers. Second, various research groups have resequenced additional sorghum accessions for SNP discovery. There is no foreseeable obstacle to obtaining hundreds of thousands of SNPs for genome-wide scans for multiple traits. Genotyping arrays with different marker densities have been developed, and the density is expected to increase. In addition, genotyping-by-sequencing may soon become practical for these sorghum diversity panels. Third, our understanding of association mapping panels and analysis methods has significantly increased because of earlier empirical studies in LD and association mapping.

Phenotyping, however, remains a major challenge, especially for agronomically important traits. Obtaining robust phenotypes (e.g., abiotic and biotic stresses) for a large number of accessions requires multiple environmental trials, long-term commitment, and stable funding for a concerted research consortium. Association mapping is multidisciplinary in nature and could be difficult to implement in small research programs; many aspects of this approach deserve further attention.

Fortunately, preliminary efforts have been made to address the adequacy of background markers for estimating population structure and relative kinship

(Yu et al. 2009), variation explained with mixed-model association mapping (Sun et al. 2010), and computational efficiency in large-scale, genome-wide studies (Zhang et al. 2010). As we move toward GWAS, the question of missing heritability may emerge. Quantitative genetics has played a critical role in developing plant and animal breeding methods and provides a natural framework for dissecting complex traits with high-throughput technologies. Modifying and adapting classic quantitative genetics and population genetics models and combining genomic technologies with genetic designs and experimental designs will help us find those missed heritabilities. Ultimately, association genetics is an additional strategy that needs to be combined with existing and emerging strategies to realize the full potential of ultra-high-throughput genomic technologies in crop improvement (Yu 2009).

Acknowledgments This project is supported by the Kansas State University Targeted Excellence Program, National Research Initiative (NRI) Plant Genome Program of the USDA Cooperative State Research, Education and Extension Service (CSREES) (2006-35300-17155), Kansas Grain Sorghum Commission, International Sorghum and Millet Collaborative Research Support Program (INTSORMIL-CRSP), and Institute for Genomic Diversity at Cornell University.

References

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408:796–815

Atwell S, Huang YS, Vilhjalmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT, Jiang R, Muliyati NW, Zhang X, Amer MA, Baxter I, Brachi B, Chory J, Dean C, Debieu M, de Meaux J, Ecker JR, Faure N, Kniskern JM, Jones JD, Michael T, Nemri A, Roux F, Salt DE, Tang C, Todesco M, Traw MB, Weigel D, Marjoram P, Borevitz JO, Bergelson J, Nordborg M (2010) Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature 465:627–631

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265

Bernai JH, Clark RB (1998) Growth traits among sorghum genotypes in response to aluminium. J Plant Nutr 21:297–305

Bernardo R (2008) Molecular markers and selection for complex traits in plants: learning from the last 20 years. Crop Sci 48:1649–1664

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635

Brown PJ, Rooney WL, Franks C, Kresovich S (2008) Efficient mapping of plant height quantitative trait loci in a sorghum association population with introgressed dwarfing genes. Genetics 180:629–637

Bucheyeki TL, Shenkalwa EM, Mapunda TX, Matata LW (2010) Yield performance and adaptation of four sorghum cultivars in Igunga and Nzega districts of Tanzania. Commun Biometry Crop Sci 5:4–10

Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC, Goodman MM, Harjes C, Guill K, Kroon DE, Larsson S, Lepak NK, Li H, Mitchell SE, Pressoir G, Peiffer JA, Rosas MO, Rocheford TR, Romay MC, Romero S, Salvo S, Sanchez Villeda H, da Silva HS, Sun Q, Tian F, Upadyayula N, Ware D, Yates H, Yu J, Zhang Z, Kresovich S, McMullen MD (2009) The genetic architecture of maize flowering time. Science 325:714–718

Carpita NC, McCann MC (2008) Maize and sorghum: genetic resources for bioenergy grasses. Trends Plant Sci 13:415–420

Casa AM, Pressoira G, Brown PJ, Mitchell SE, Rooney WL, Tuinstrac MR, Franks CD, Kresovicha S (2008) Community resources and strategies for association mapping in sorghum. Crop Sci 48:30–40

Casas AM, Kononowicz AK, Zehr UB, Tomes DT, Axtell JD, Butler LG, Bressan RA, Hasegawa PM (1993) Transgenic sorghum plants via microprojectile bombardment. Proc Natl Acad Sci U S A 90:11212–11216

de Alencar Figueiredo LF, Calatayud C, Dupuits C, Billot C, Rami JF, Brunel D, Perrier X, Courtois B, Deu M, Glaszmann JC (2008) Phylogeographic evidence of crop neodiversity in sorghum. Genetics 179:997–1008

Deu M, Rattunde F, Chantereau J (2006) A global view of genetic diversity in cultivated sorghums using a core collection. Genome 49:168–180

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357–374

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchison D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Paszkowski U, Zhang S, Colbert M, Sun WL, Chen L, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu Y, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T, Oliphant A, Briggs S (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). Science 296:92–100

Gowda CLL, Serraj R, Srinivasan G, Chauhan YS, Reddy BVS, Rai KN, Nigam SN, Gaur PM, Reddy LJ, Dwivedi SL, Upadhyaya HD, Zaidi PH, Rai HK, Maniselvan P, Folkerstma R, Nalini M (2009) Opportunities for improving crop water productivity through genetic enhancement of dryland crops. In: Wani SP, Rockström J, Oweis T (eds) Rainfed agriculture: unlocking the potential. CAB International, Wallingford, pp 133–163

Hamblin MT, Casa AM, Sun H, Murray SC, Paterson AH, Aquadro CF, Kresovich S (2006) Challenges of detecting directional selection after a bottleneck: lessons from Sorghum bicolor. Genetics 173:953–964

Hamblin MT, Fernandez MGS, Tunistra MR, Rooney WL, Kresovich S (2007) Sequence variation at candidate loci in the starch metabolism pathway in sorghum: prospects for linkage disequilibrium mapping. Crop Sci 47:S125–S134

Hamblin MT, Mitchell SE, White GM, Gallego J, Kukatla R, Wing RA, Paterson AH, Kresovich S (2004) Comparative population genetics of the panicoid grasses: sequence polymorphism, linkage disequilibrium and selection in a diverse sample of sorghum bicolor. Genetics 167:471–483

Hamblin MT, Salas Fernandez MG, Casa AM, Mitchell SE, Paterson AH, Kresovich S (2005) Equilibrium processes cannot explain high levels of short- and medium-range linkage disequilibrium in the domesticated grass Sorghum bicolor. Genetics 171:1247–1256

Harjes CE, Rocheford TR, Bai L, Brutnell TP, Kandianis CB, Sowinski SG, Stapleton AE, Vallabhaneni R, Williams M, Wurtzel ET, Yan J, Buckler ES (2008) Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. Science 319:330–333

Holland JB (2007) Genetic architecture of complex traits in plants. Curr Opin Plant Biol 10: 156–161

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM (2009) Finding the missing heritability of complex diseases. Nature 461:747–753

McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q, Flint-Garcia S, Thornsberry J, Acharya C, Bottoms C, Brown P, Browne C, Eller M, Guill K, Harjes C, Kroon D, Lepak N,

Mitchell SE, Peterson B, Pressoir G, Romero S, Oropeza Rosas M, Salvo S, Yates H, Hanson M, Jones E, Smith S, Glaubitz JC, Goodman M, Ware D, Holland JB, Buckler ES (2009) Genetic properties of the maize nested association mapping population. Science 325:737–740

Miller FR (1984) Registration of RTx430 sorghum parental line. Crop Sci 24:1224

Murray SC, Rooney WL, Hamblin MT, Mitchell SE, Kresovich S (2009) Sweet sorghum genetic diversity and association mapping for brix and height. Plant Genome 2:48–62

Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES (2009) Association mapping: critical considerations shift from genotyping to experimental design. Plant Cell 21: 2194–2202

Nordborg M, Tavare S (2002) Linkage disequilibrium: what history has to tell us. Trends Genet 18: 83–90

Nordborg M, Weigel D (2008) Next-generation genetics in plants. Nature 456:720–723

Paterson AH, Bowers JE, Bruggmann R, Dubchak I et al (2009) The Sorghum bicolor genome and the diversification of grasses. Nature 457:551–556

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273:1516–1517

Rooney WL, Blumenthal J, Bean B, Mullet JE (2007) Designing sorghum as a dedicated bioenergy feedstock. Biofuels Bioprod Bioref 1:147–157

Schmutz J, Cannon SB, Schlueter J, Ma J et al (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183

Schnable PS, Ware D, Fulton RS, Stein JC et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115

Shehzad T, Iwata H, Okuno K (2009a) Genome-wide association mapping of quantitative traits in sorghum (Sorghum bicolor (L.) Moench) by using multiple models. Breed Sci 59:217–227

Shehzad T, Okuizumi H, Kawase M, Okuno K (2009b) Development of SSR-based sorghum (Sorghum bicolor (L.) Moench) diversity research set of germplasm and its evaluation by morphological traits. Genet Resour Crop Evol 56:809–827

Sorrells ME, Yu J (2009) Linkage disequilibrium and association mapping in the Triticeae. In: Feuillet C, Muehlbauer GJ (eds) Genetics and genomics of the Triticeae, plant genetics/genomics. Springer, New York, pp 655–684

Sun G, Zhu C, Kramer MH, Yang SS, Song W, Piepho HP, Yu J (2010) Variation explained in mixed-model association mapping. Heredity 105:333–340

Tian Z, Qian Q, Liu Q, Yan M, Liu X, Yan C, Liu G, Gao Z, Tang S, Zeng D, Wang Y, Yu J, Gu M, Li J (2009) Allelic diversities in rice starch biosynthesis lead to a diverse array of rice eating and cooking qualities. Proc Natl Acad Sci U S A 106:21760–21765

Upadhyaya HD, Pundir RPS, Dwivedi SL, Gowda CLL, Reddy VG, Singh S (2009) Developing a mini core collection of sorghum for diversified utilization of germplasm. Crop Sci 49: 1769–1780

Visscher PM (2008) Sizing up human height variation. Nat Genet 40:489–490

Wang ML, Zhu C, Barkley NA, Chen Z, Erpelding JE, Murray SC, Tuinstra MR, Tesso T, Pederson GA, Yu J (2009) Genetic diversity and population structure analysis of accessions in the US historic sweet sorghum collection. Theor Appl Genet 120:13–23

Wang WY, Barratt BJ, Clayton DG, Todd JA (2005) Genome-wide association studies: theoretical and practical concerns. Nat Rev Genet 6:109–118

Yan J, Kandianis CB, Harjes CE, Bai L, Kim EH, Yang X, Skinner DJ, Fu Z, Mitchell S, Li Q, Fernandez MG, Zaharieva M, Babu R, Fu Y, Palacios N, Li J, Dellapenna D, Brutnell T, Buckler ES, Warburton ML, Rocheford T (2010) Rare genetic variation at Zea mays crtRB1 increases beta-carotene in maize grain. Nat Genet 42:322–327

Yu J (2009) The potential of ultrahigh throughput genomic technologies in crop improvement. Plant Genome 2:2

Yu J, Buckler ES (2006) Genetic association mapping and genome organization of maize. Curr Opin Biotechnol 17:155–160

Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. Genetics 178:539–551

Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Huang X, Li W, Li J, Liu Z, Li L, Liu J, Qi Q, Liu J, Li L, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Zhang J, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Ren X, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Wang J, Zhao W, Li P, Chen W, Wang X, Zhang Y, Hu J, Wang J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Li G, Liu S, Tao M, Wang J, Zhu L, Yuan L, Yang H (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. indica). Science 296:79–92

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208

Yu J, Zhang Z, Zhu C, Tabanao DA, Pressoir G, Tuinstra MR, Kresovich S, Todhunter RJ, Buckler ES (2009) Simulation appraisal of the adequacy of number of background markers for relationship estimation in association mapping. Plant Genome 2:63–77

Zhang Z, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM, Buckler ES (2010) Mixed linear model approach adapted for genome-wide association studies. Nat Genet 42:355–360

Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, Toomajian C, Zheng H, Dean C, Marjoram P, Nordborg M (2007) An Arabidopsis example of association mapping in structured samples. PLoS Genet 3:e4

Zhao ZY, Cai T, Tagliani L, Miller M, Wang N, Pang H, Rudert M, Schroeder S, Hondred D, Seltzer J, Pierce D (2000) Agrobacterium-mediated sorghum transformation. Plant Mol Biol 44:789–798

Zhu C, Gore MA, Buckler ES, Yu J (2008) Status and prospects of association mapping in plants. Plant Genome 1:5–20

Zhu C, Yu J (2009) Nonmetric multidimensional scaling corrects for population structure in whole genome association studies. Genetics 182:875–888

Chapter 10
Sorghum Transformation: Overview and Utility

Tejinder Kumar, Arlene Howe, Shirley Sato, Ismail Dweikat, and Tom Clemente

Abstract Over the past decade genomics resources available for sorghum have rapidly expanded (Paterson Int J Plant Genomics 2008:6, 2008), these resources, coupled with the recent completion of the genome sequence which is relatively small in size (730 Mb) (Paterson et al. Nature 457:551–556, 2009) makes sorghum a rather attractive species to study. Moreover, the USDA germplasm system maintains 42,614 accessions, of which more than 800 exotic landraces have been converted to day length-insensitive lines to facilitate their use in breeding programs. In addition, a set of EMS mutation stocks developed by the USDA Plant Stress and Germplasm Development Unit in Lubbock, TX (Xin et al. Bioenerg Res 2:10–16, 2009) will be a valuable resource for functional genomics studies in sorghum.

T. Kumar (✉) • I. Dweikat
Department of Agronomy & Horticulture, University of Nebraska-Lincoln,
1400 R Street, Lincoln, NE 68583, USA
e-mail: tejinderkumar@rediffmail.com

A. Howe
Center for Biotechnology, University of Nebraska-Lincoln, Lincoln, NE 68588, USA

Monsanto Company, Chesterfield, MO 63017, USA

S. Sato
Center for Biotechnology, University of Nebraska-Lincoln,
Lincoln, NE 68588, USA

T. Clemente
Department of Agronomy & Horticulture, University of Nebraska-Lincoln,
1400 R Street, Lincoln, NE 68583, USA

Center for Biotechnology, University of Nebraska-Lincoln,
Lincoln, NE 68588, USA

Center for Plant Science Innovation, University of Nebraska-Lincoln,
Lincoln, NE 68588, USA

A.H. Paterson (ed.), *Genomics of the Saccharinae*, Plant Genetics and Genomics: Crops and Models 11, DOI 10.1007/978-1-4419-5947-8_10,

However, in order to be a robust system for study a suite of functional genomics tools are necessary to complement these other resources to aid in down-stream hypothesis testing. A key functional genomics tool is the ability to modulate gene expression through the introduction of transgenic genetic elements. This is exemplified by recent work (Cook et al. Plant Cell 22:867–887, 2010) in which RNAi experiments were employed to specifically reduced expression of two alkylresorcinol synthases to demonstrate their role in the synthesis of the allelopathic molecule sorgoleone. In addition to its value as a functional genomics tool, plant transformation offers a route to broaden access to novel input and output traits for sorghum breeding programs.

Keywords *Agrobacterium tumefaciens* • Transformation • *npt II* • Biotechnology • Sorghum • Genetic engineering

1 Sorghum Transformation

In general plant transformation can be partitioned into two components: competence of a cell for culture regeneration into a whole plant and receptiveness of that same cell for foreign DNA integration. In sorghum, like most monocotyledonous plants, in vitro culture regimes are primarily somatic embryogenesis based systems (Elkonin and Pakhomova 2000; Jogeswar et al. 2007; Kaeppler and Pedersen 1996; Pola et al. 2008; Pola and Mani 2006; Sato et al. 2004a). As per the second component of plant transformation, integration of genetic elements, sorghum has been successfully transformed using both direct DNA delivery methods (Battraw and Hall 1991) and *Agrobacterium*-mediated transformation protocols (Cai et al. 2002; Gao et al. 2005a, b; Gurel et al. 2009; Howe et al. 2006; Nguyen et al. 2007; Zhao et al. 2000). While both DNA delivery systems are proven technologies for recovery of stable sorghum transformants, more laboratories are moving towards implementing the latter due to the tendency of *Agrobacterium*-mediated transformants to carry lower copy number insertions and/or have a higher frequency of coexpression of the nonselected transgenic cassette (Dai et al. 2001; Gao et al. 2008; Zhao et al. 1998).

While multiple explants have been evaluated as the starting material for sorghum transformation, clearly the primary explant reported on is immature embryos. One of the factors that have hampered transformation efficiencies of sorghum with the immature embryo explant is the rapid production of phenolic compounds. Phenolics are produced during the in vitro culturing of sorghum immature embryos, but the production of these secondary metabolites is enhanced upon inoculation with *A. tumefaciens*. To alleviate the negative effects of phenolics on sorghum transformation media supplements such as polyvinylpolypyrrolidone (PVPP) (Cai et al. 1987), and elevation of potassium phosphate levels (Elkonin and Pakhomova 2000; Sato et al. 2004a), or the exposure of explants to reduced temperature (Nguyen et al. 2007) have been shown to be able to reduce, but not totally eliminate the negative impact of these compounds. Triggering of the plant's defense response upon

challenge with *A. tumefaciens* may lead not only to the production of secondary metabolites, but also to cell death, which can further hamper the efficiency of recovery of transgenic plants. For example in banana the triggering of apoptosis by *A. tumefaciens* can be effectively countered by the expression of antiapoptotic genes (Khanna et al. 2007). While such a strategy has not been evaluated in sorghum, a heat shock pretreatment, which was previously reported to counter apoptosis in banana embryogenic callus, leading to improved transformation efficiency (Khanna et al. 2004), has recently been shown to be a translatable technique using sorghum immature embryos (Gurel et al. 2009).

Key to any transformation system is the ability to rapidly, and efficiently distinguish transgenic differentiating cells from nontransgenic cell lineages. Two means typically used to differentiate transgenic from nontransgenic cell lineages are the use of visual or selectable marker genes. In sorghum the visual marker genes green fluorescent protein (gfp) and ß-glucuronidase (GUS) are each effective in monitoring for transgenic cells (Jeoung et al. 2002). Using the former visual marker Gao et al. (2005a) reported a 3.0% transformation efficiency as means to monitor for transgenic differentiating cell lineages from immature embryos of sorghum.

Selectable marker genes used to provide a competitive edge in culture for plant transformation systems typically rely upon providing resistance to antibiotics, such as hygromycin (Gritz and Davies 1983), and the aminoglycoside kanamycin, or various derivatives thereof (Fraley et al. 1983) or tolerance towards herbicidal agents glyphosate (Barry et al. 1992) and glufosinate (Thompson et al. 1987). In addition the positive selectable marker gene phosphomannose isomerase (PMI) (Joersbo and Okkels 1996) has been shown to be a rather robust selection system for the identification of transgenic plants (Negrotto et al. 2000), including sorghum (Gao et al. 2005b).

1.1 Outline of an Agrobacterium-Mediated Transformation of Sorghum Using npt II as a Selectable Marker Gene

As indicated above there have been multiple reports of successful transformation of sorghum following the communicated success in 1993 (Casas et al. 1993). Outlined below is the system reported on by Howe et al. (2006) that utilizes *npt II* as the selectable marker gene, coupled with G418 a the selection agent. While the overall transformation efficiency with this system is relatively low, typically ranging from 1 to 3%, the system is consistent, and importantly teachable with minimal training.

This sorghum transformation system relies upon immature embryos as the starting material. One of the disadvantages of using this explant is the need for continual plantings of stock plants to ensure a constant supply of immature embryos, adding labor and cost to the system. Nonetheless, the stock plants used to supply immature embryos are maintained under greenhouse conditions. Heads are harvested when 70% of the head have embryos ranging in size from 1.2 to 2.2 mm in length. Each head is then excised from the plant and placed in a 1,000 ml graduated cylinder

filled with 500 ml of 50% commercial bleach plus 250 ml of Tween 20. The cylinder containing the sorghum head submerged in the bleach solution is mixed with a stir bar for 20 min. Following the 20 min surface sterilization with bleach the heads are washed three times with sterile water and allowed to air dry in laminar flow hood. Seeds are isolated and subjected to a secondary sterilization process consisting of a 1 min soak with agitation in 70% ethanol, followed by a single wash with sterile water, proceeded by a soak with agitation in 10% bleach solution, with a subsequent triple wash in sterile water.

Immature embryos are isolated from 50 sterilized seeds and placed in a 35 × 10 mm Petri plate containing 1 ml of coculture medium supplemented with 300 μM of acetosyringone. The coculture medium is composed of 0.5× MS major and minor salts (Murashige and Skoog 1962), 0.5 mg/l each of nicotinic acid and pyridoxine HCl, plus 1 mg/l each of thiamine HCl and casamino acids. The carbohydrate sources are 2% sucrose and 1% glucose and growth regulator, 2,4-D, at 2 mg/l. The medium is buffered with 3 mM MES (pH 5.2). The medium is filter sterilized, with no components being autoclaved.

Once 50 immature embryos have been isolated the coculture medium is removed and replaced with 1 ml of *A. tumefaciens* inoculum. The inoculum is an *A. tumefaciens* strain NTL_4/pTiPKPSF2 (Palanichelvam et al. 2000), suspension in cocultivation medium (OD_{660} 0.3–0.5). Inoculation time is 5 min. Following the cocultivation step the explants are placed scutellum side up on 100 × 20 mm Petri plates containing four sterile Whatman™ filter papers saturated with 4.2 ml of cocultivation medium. The plates are incubated for 2 days at 24°C in the dark.

Following the cocultivation period the explants are cultured on delay medium which is composed of Elkonin's major salts (Elkonin and Pakhomova 2000), MS minor salts and vitamin mix, 2 g/l proline, and 1 g/l asparagine. The carbohydrate source is 3% sucrose, the medium is buffered with 3 mM MES (pH 5.7) and solidified with 2% phytagel. To counter select against *A. tumefcaiens* the medium is supplemented with 100 mg/l carbenicillin. The growth regulator 2,4-D is used at a level of 1.5 mg/l. Cultures are incubated at 28°C in the dark for 3 days.

The selection phase is immediately implemented following the delay period. A total of 20 embryos are placed on to 100 × 20 mm Petri plates containing the delay medium supplemented with 20 mg/l G418. The tissue is transferred to fresh selection medium every 2–3 weeks. As coleoptiles develop they are systematically removed from the explants. As the embryogenic callus begins to form about the explants the tissue is broken up into 1–3 mm pieces, with care being taken to ensure tracking of tissue from the original explant, given the fact that most transformants derived from the same initial immature embryo tend to be clones, and hence the best way to track transformation efficiencies. The callus tissue remains in the selection phase for a period of 6–9 weeks.

Following the selection phase proliferating embryogenic tissue is transferred to regeneration medium composed of MS major/minor salts and vitamins, supplemented with 0.5 mg/l kinetin and 1.0 mg/l IAA. The medium is solidified with 2% phytagel, carbohydrate level, 3% sucrose, and buffered with 3 mM MES (pH 5.7). The selection pressure is reduced to 10 mg/l G418, and the carbenicillin level

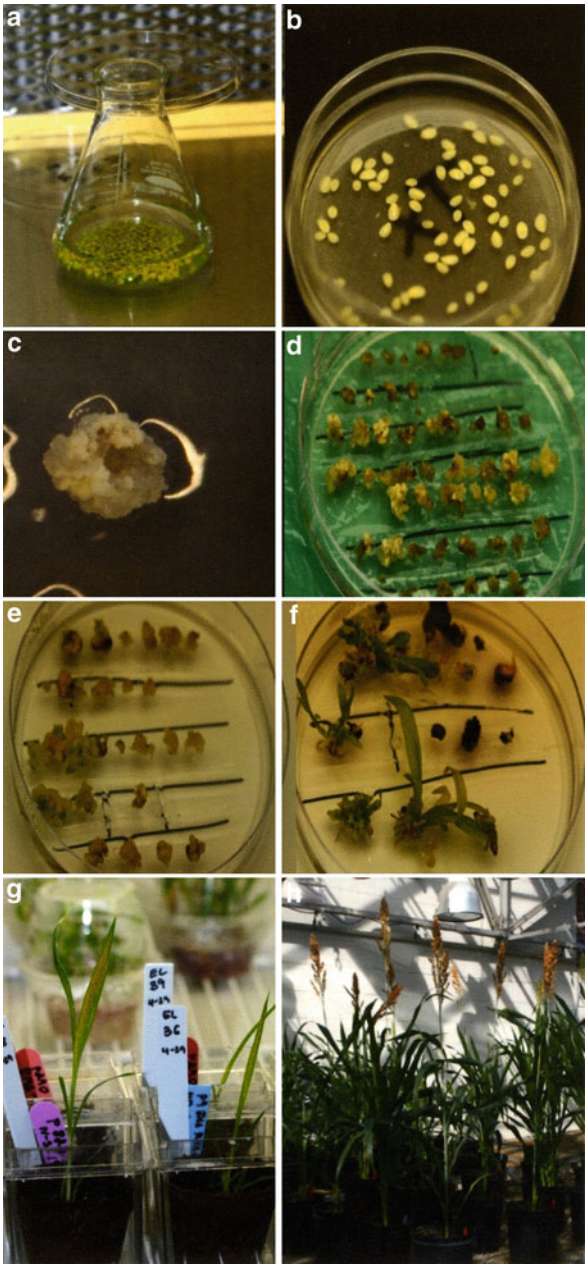

Fig. 10.1 Overview of sorghum transformation steps. (**a**) Sterilization step of immature seeds. (**b**) Inoculation step of immature embryos. (**c**) Somatic embryogeneic tissue. (**d**) Partitioning of somatic embryogenic tissue to ensure lineage tracking. (**e**, **f**) Regeneration steps. (**g**) Acclimation of plants. (**h**) Transgenic sorghum in greenhouse

remains at 100 mg/l. The cultures are placed under a 16/8 light regime at 24°C. Typically after 4 weeks on regeneration, with one subculture at the 2-week period, shoots with well established roots will form, that are ready to be acclimated to soil (Fig. 10.1).

Once acclimated the primary transformants are confirmed using a commercial ELISA kit to monitor *npt II* expression (Agdia Corp.). Primary transformants are screened to identify a minimum of two lead events for down-stream characterizations. A lead event is selected based on simple integration pattern of the transgenic element(s), and expression of the target phenotype(s) of interest.

2 Considerations in Designing Binary Vectors for Sorghum Transformation

An attribute of *Agrobacterium*-mediated transformation is that T-DNAs can be integrated, albeit relatively infrequently, at unlinked positions. This ability of *A. tumefaciens* can be exploited to derived marker-free transgenic events through the simultaneous delivery of two T-DNAs, where one of the T-DNA elements carries the marker gene, and the other carries gene(s) of interest. If integrated at unlinked positions, the T-DNAs, will segregate in the progeny. This strategy has been successfully used to derived maker-free transgenic plants in a number of systems (Daley et al. 1998; Jacob and Veluthambi 2002; Komari et al. 1996; Sato et al. 2004b; Xing et al. 2000), including sorghum (Zhao et al. 2003). The integration of unlinked T-DNA alleles in sorghum is exemplified in Fig. 10.2. Transgenic sorghum events were generated that harbored a transgenic cassette with the cyanamide hydratase (cah) gene (Maier-Greiner et al. 1991), under control of the sugarcane polyubiquitin promoter ubi4 (Wei et al. 2003), housed within a single T-DNA binary vector designated pPTN181 (not shown). A Southern blot analysis is performed using a restriction enzyme wherein one recognition site resides within the T-DNA element, hence each hybridization signal will reflect a single integration locus, on a subset of primary transformants derived from pPTN181 as shown in Fig. 10.2a. As can be seen the event 168 carries one locus, while events 165 and 166 harbor two loci, and event 164 contains three loci. Monitoring segregation of the transgenic alleles in progeny derived from these events revealed a 15:1 pattern for events 165 and 166, and a 3:1 pattern for events 164 and 168 (data not shown). Southern analysis on a subset of the derived progeny is in agreement with the observed segregation patterns (Fig. 10.2b, c). It can be seen in the T_1 individuals derived from events 165 and 166 that some individuals display the genotype of the parent, while others only carry one of the transgenic alleles. On the other hand all T_1 derived from event 164 had the same genotype as the parent, hence all alleles appear to be linked. However the single locus event, 168, segregated as expected, with T_1 individuals genotyped the same as the parent.

When implementing the tool of plant transformation for targeted output and input traits in sorghum such as improvement in grain quality or stress tolerance, respective, it is critical to have promoter elements with the desired specificity so to limit the probability of negatively impacting agronomic performance that may arise if the phenotype is misexpressed in nontarget tissues. To this end it is prudent to verify promoter specificity if using a promoter known to be tissue specific in other

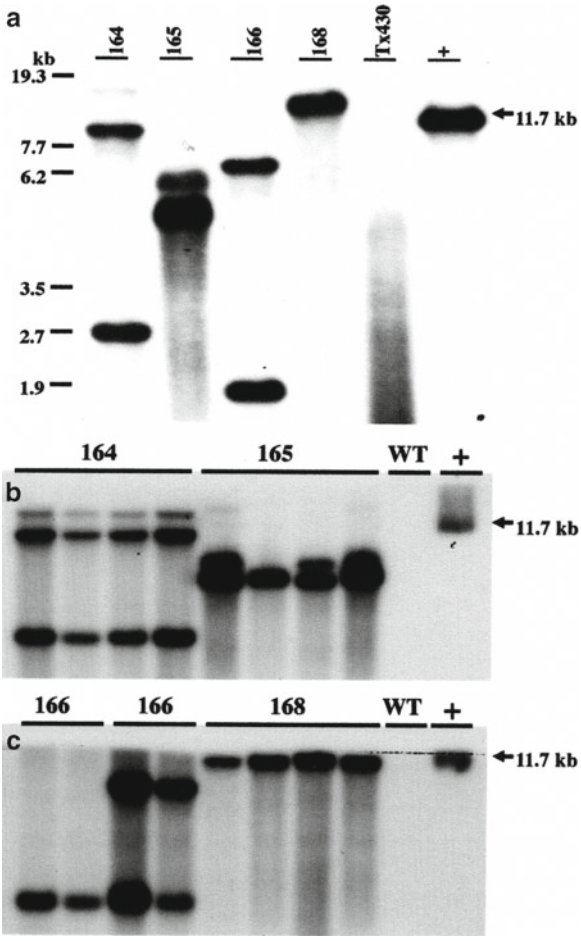

Fig. 10.2 Southern blot analysis of transgenic sorghum events carrying cyanamide hydratase (cah) gene. (**a**) Primary transformation events designated 164, 165, 166, and 168 probed with cah ORF. + lane indicates 50 pg linear binary vector pPTN181. Tx430 lane is 10 μg of wild type DNA. (**b**) Southern blot analysis of T_1 progeny derived from events 164 and 165, highlighting segregating transgenic alleles in event 165, and linked alleles in 164. *WT* lane indicates 10 μg genomic DNA from Tx430. + lane is 50 pg of linear binary vector pPTN181. (**c**) Southern blot analysis of T_1 progeny derived from events 166 and 168, highlighting segregating transgenic alleles in event 166, and linked locus in 168. *WT* lane indicates 10 μg genomic DNA from Tx430. + lane is 50 pg of linear binary vector pPTN181

species, before assembling cassettes for use in sorghum. For example interest in modifying seed components of sorghum may require specific expression in the embryo. A logical candidate promoter for desired embryo-specific expression would be the maize globulin-1 promoter (Belanger and Kriz 1991). To evaluate whether the glob-1 promoter specificity would translate to sorghum a GUSPlus™ (www.cambia.org) cassette under control of the glob-1 promoter was assembled and introduced into sorghum. As a constitutive control transgenic sorghum carrying a GUSPlus™ cassette under control of the maize polyubiquitin promoter (Christensen et al. 1992) was used for comparison. Tissue samples were assayed over development in T_1 or T_2 individuals, looking at GUS expression within root, leaf, stem, glume, scutellum, and embryos. As can be seen in Fig. 10.3, embryo-specificity of the maize glob-1 promoter does effectively translate to sorghum. While this result is not surprising, these are data that need to be gathered to fully exploit sorghum transformation as a translational genomics tool.

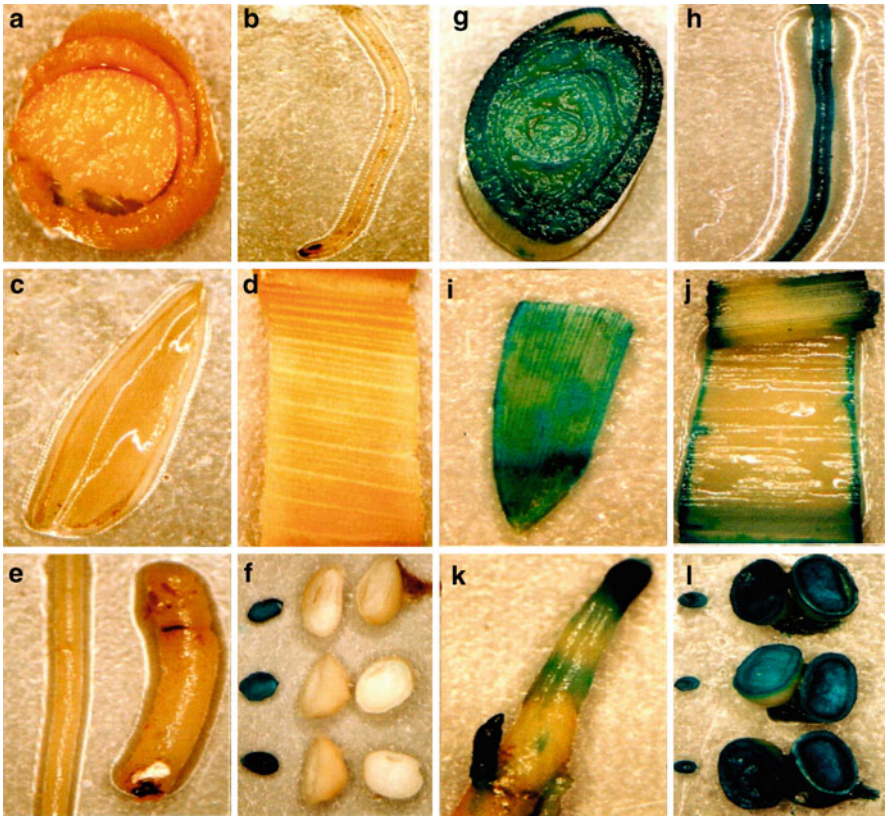

Fig. 10.3 GUS expression profile observed in transgenic sorghum. Panels (**a**–**f**): Transgenic sorghum event carrying the glob-1-GUS cassette showing embryo specific expression. (**a**) Stem section, (**b**) root section, (**c**) glume, (**d**) leaf, (**e**) root, and (**f**) seed, endosperm and embryo (*blue*). Panels (**g**–**l**): Transgenic sorghum event carrying the ubiquitin-1-GUS cassette showing constitutive GUS expression. (**g**) Stem section, (**h**) root section, (**i**) glume, (**j**) leaf, (**k**) root, and (**l**) seed, endosperm and embryo

3 Target Input Traits for Sorghum Through Transformation

A critical trait for any breeding program is yield. Addressing yield directly through transgenic approaches is a rather large challenge. A more practical and obtainable goal in the short term is protection of yield through control of biotic and abiotic stresses. In sorghum, like most crops, key stresses that compromise yield will vary across regions. Sorghum production can be severely impacted by a number of insect pests. Not only can insects impact production directly, but in some cases they can also provide an entry for secondary pathogen attack at the site of insect feeding. The success of the *Bt* technology in maize (Armstrong et al. 1995; Barry et al. 2000), and cotton (Cattaneo et al. 2006), is a strong rationale for the evaluation of this

technology in sorghum as a means to combat specific target insect pests. Importantly, in addition to *Bt*'s direct impact in impeding insect pest feeding damage, a secondary effect observed with the use of this technology is a significant reduction in accumulation of various mycotoxins in plant tissues (Abbas et al. 2008; Bakan et al. 2002; Hammond et al. 2004). This secondary attribute of the *Bt* technology may serve as a valuable mechanism to limit quality issues of sorghum related to contamination of these toxins that may occur under certain conditions and fungal infestation levels (Ghali et al. 2009; Reddy and Raghavender 2008; Reddy et al. 2010). However, like all disease resistance traits, the *Bt* technology needs to be used in conjunction with proper integrated pest management practices to maximize its durability over time (Kumar and Pandey 2008; Sharma and Ortiz 2000).

A number of viral agents have been shown to be capable of replication in sorghum (Jensen and Giorda 2002), including members of the potyvirus family including sugarcane mosaic virus, maize dwarf mosaic virus, and sorghum mosaic virus. Limited resistance towards these viral agents has been identified within sorghum germplasm, although some reports have been communicated (Henzell et al. 1982). The seminal work which demonstrated introduction of viral coat protein genes in transgenic plants to confer virus resistance (Abel et al. 1986; Nelson et al. 1987; Stark and Beachy 1989), has opened the door for the translation of this technology to other plant systems, implementing various genetic constructs that target silencing of critical gene products required for the replication of the virus of interest (Beachy et al. 2003; Prins 2003), including known pathogens of sorghum (Gilbert et al. 2005). Hence, such strategies offer great potential for the introduction of durable virus resistance for sorghum.

Striga, commonly referred to as witch weed, contains two species, *S. hermonthica* and *S. asiatica*, that are parasitic on sorghum and other cereals (Aly 2007). Parasitic plant species infest nearly 50 million hectares crop plants on an annual basis, and great strides have been made in developing resistance in sorghum through conventional breeding approaches (Ejeta 2007). More recently Tuinstra et al. (2009) have communicated a herbicide seed treatment strategy that exploits the introgression of acetolactate synthase (ALS) herbicide resistance from shattercane into elite sorghum genotypes (Hennigh et al. 2010). Implementing this seed-coating approach significantly reduced *Striga* emergence under both greenhouse and field tests (Tuinstra et al. 2009). While this is a very promising tool to combat this devastating parasite, given that ALS inhibiting herbicides are typically classified as high risk for development of resistance, the durability of such a strategy may be limited without proper management. Hence, other approaches are needed to ensure long-term control towards *Stiga*. To this end, there has been a report looking at targeting critical genes in parasitic plant's life cycle by expression of hair-pin constructs in the host plant which resulted in an enhanced tolerance phenotype in the *Orobanche aegyptiaca*/tomato host parasite interaction (Aly et al. 2009). However, this approach was not successfully translated as a means to control to the *Striga*/maize parasite interaction (Yoder and Scholes 2010). Clearly additional research is required to further our understanding of the underlying biology involved during the early stages of parasitism by these plants. More efforts around the assembly of -omics databases

that carry this information (Torres et al. 2005) are needed in order to facilitate the development of alternative control strategies towards *Striga*, that may serve as a complement to the herbicide seed coating approach (Tuinstra et al. 2009).

Addressing a plant's response to stresses that are governed in a multigenic fashion is more challenging than single gene traits. In order to investigate multigenic abiotic stress response traits such as drought and heat, researchers are evaluating a coordinated expression of a suite of genes triggered by exposure to the targeted stress by the introduction of a single transcription factor (Karaba et al. 2007; Nelson et al. 2007; Suzuki et al. 2005). Theses transcription-factor based technologies hold great promise as a means to reduce multigenic expressed phenotypes to a single transgene fashion (Century et al. 2008), however, the transcription factor based strategy undoubtedly will require tight regulation, necessitating the need for tissue-specific and/or inducible promoter systems.

With respect to adaptation to low nitrogen environments, Yanagisawa et al. (2004) demonstrated that expression of the maize *Dof1* transcription factor improved nitrogen assimilation in transgenic plants. However, it is feasible to directly perturb nitrogen flux in plants. Nitrogen assimilation and metabolism in plants occurs through coordinated action of a variety of enzymes acting upon a variety of substrates. Two key enzymes involved in nitrogen metabolism in plants are glutamine synthetase (GS) and glutamate synthase (GOGAT). Previous studies have shown that enhancing GS or GOGAT activities can impact nitrogen metabolism in plant species (Cai et al. 2009; Good et al. 2004). Enhancing activity of another enzyme that impacts nitrogen flux in plants, alanine aminotransferease (Ala-AT), that catalyzes the production of alanine and 2-oxoglutarate from pyruvate and glutamate, has been shown to augment nitrogen use efficiency in both rape seed and rice (Good et al. 2007; Shrawat et al. 2008).

A caveat to these studies communicating enhancing nitrogen use efficiency through transgenic approaches is that most reports used data sets gathered from greenhouse or growth chamber studies, with minimal information on the impact of the respective transgenes on yield under field conditions (Brauer and Shelp 2010). Moreover, no data sets have been communicated to date on the impact of gene stacking strategies on nitrogen use efficiencies with these selected genes.

4 Target Output Traits for Sorghum Through Transformation

Digestibility of sorghum limits protein availability, and ultimately impacts the nutritional quality of the grain (Duodu et al. 2003). The major proteins, prolamins, found in sorghum reside in the endosperm. The prolamin storage proteins found in sorghum and maize are designated kafirins and zeins, respectively. The prolamins are assembled into protein bodies, with a very defined pattern, where the α class reside in the core along with the δ class, albeit to a lower extent, while the β and γ classes decorate the periphery of the protein body (Wu and Messing 2010). A number of parameters can influence digestibility of sorghum protein, including structure and

shape of the protein body (Duodu et al. 2003). Reduction of the zein proteins found in the maize mutants *opaque-2* (Hartings et al. 1989) and *floury-2* (Coleman et al. 1995) leads to a concomitant increase in lysine and tryptophan due to a compensation mechanism in seeds resulting in an increase in nonzein proteins (Coleman and Larkins 1999). Deliberate reduction in the 19 kDa α-zeins in maize manifests the opaque kernel phenotype, and enhances levels of lysine and tryptophan in the grain (Huang et al. 2004). Similarly, reduction in the level of both the β- and γ-zeins resulted in drastic changes protein bodies, and triggered the opaque kernel phenotype (Wu and Messing 2010). Hence, modulation of the prolamins is a target that could be pursued in sorghum as a means to simultaneously address digestibility, and nutritional quality.

Oria et al. (2000) described a highly digestible, enhanced lysine sorghum mutant. The protein bodies observed within this mutant are highly folded, with a redistribution of the γ-kafirin about the body. These factors lead to increased exposure of the core α-kafirins, which translates to the increased digestibility phenotype (Duodu et al. 2003).

Like the maize *floury-2* and *opaque-2* mutants, the highly digestible, enhanced lysine mutant of sorghum has value in both food and feed applications. However, translation of these traits to application has yet to be realized, undoubtedly due to the tendency of these altered prolamin grains to have reduced agronomic properties, and post harvest issues (Huang et al. 2004). However, these drawbacks may not be insurmountable. Breeding efforts are making progress in addressing the issues blocking deployment of the high digestible, enhanced lysine mutant of sorghum (Tesso et al. 2006, 2008a). Through better understanding of the underlying biology governing protein deposition in these mutants and the influence of the various genetic modifiers, will lead to improved biotechnology approaches, coupled with better breeding strategies, to modulate the seed storage proteins, without negatively altering the endosperm characteristics. Hence, in the end, the successful deployment of a high quality grain sorghum will ultimately require an interdisciplinary approach that brings together the expertise of plant breeding, biotechnology, molecular biology/genetics and food science.

5 Potential of Outcrossing

One of the concerns raised about release of transgenic sorghum is the potential for outcrossing to its weedy relatives, johnsongrass and shattercane, which has been hypothesized to potentially lead to altered balance in the ecosystem, changes in the plant community structure, and persistence of weeds in agricultural lands (Tesso et al. 2008b).

A number of parameters must be met for a successful sorghum outcross event to occur. First, the crop and weed species must be in close proximity, and flowering times synchronized. Hybrids derived from outcrosses between grain sorghum and shattercane (*S. bicolor* subsp. *drummondii* Nees ex Steud de Wet & Harlan) do not

appear to be compromised in fitness (Sahoo et al. 2010). This lack of fitness drag in sorghum × shattercane hybrids has a benefit when introgressing desirable alleles from the shattercane into the cultivated genotypes, for example ALS resistance (Hennigh et al. 2010). However, this attribute that benefits conventional breeding strategies used to broaden diversity of cultivated sorghum, negatively impacts the use of transformation as a tool for introduction of novel traits into the crop, given there will undoubtedly be a call for more extensive regulatory testing addressing the potential ecological impact of a given transgenic sorghum event, which in turn will lead to higher costs and delay in release, with getting a transgenic sorghum event on the market.

One approach that may limit the concern of transgenic sorghum impacting the ecosystem above what is already occurring with production of conventional cultivars is the use of a male sterility system that may effectively limit pollen flow of transgenic sorghum under field conditions (Pedersen et al. 2003). However, such containment systems in many cases may not be required. The current regulatory system has a "one size fit all approach," in that regardless of the trait developed through transgenic approaches, a series of laborious and costly studies must be conducted. While it is very reasonable to assume that deployment of a transgenic sorghum event will eventually outcross to a wild relative, this must not be the deciding factor to block production. Rather than a one size fits all model, perhaps a more thoughtful, scientific regulatory process, in which decisions are made on a case-by-case model is more appropriate. Recently, Hokanson et al. (2010) have communicated a straightforward and scientific-fact based risk assessment process that hopefully will open the door for more dialog in this area, and ultimately will allow for advances in transgenic technologies to enter the marketplace expeditiously in a safe and effective manner.

References

Abbas HK, Accinelli C, Zablotowicz RM, Abel CA, Bruns HA, Dong Y, Shier WT (2008) Dynamics of mycotoxin and *Aspergillus flavus* levels in aging Bt and non-Bt corn residues under Mississippi no-till conditions. J Agric Food Chem 56:7578–7585

Abel PP, Nelson RS, De B, Hoffmann N, Rogers SG, Fraley RT, Beachy RN (1986) Delay of disease development in transgenic plants that express the tobacco mosaic virus coat protein gene. Science 232:738–743

Aly R (2007) Conventional and biotechnological approaches for control of parasitic weeds. In Vitro Cell Dev Biol 43:304–317

Aly R, Cholakh H, Joel DM, Leibman D, Steinitz B, Zelcer A, Naglis A, Yarden O, Gal-On A (2009) Gene silencing of mannose 6-phosphate reductase in the parasitic weed *Orobanche aegyptiaca* through the production of homologous dsRNA sequences in the host plant. Plant Biotechnol J 7:487–498

Armstrong CL, Parker GB, Pershing JC, Brown SM, Sanders PR, Duncan DR, Stone T, Dean DA, DeBoer DL, Hart J, Howe AR, Morrish FM, Pajeau ME, Peterson WL, Reich BJ, Rodriguez R, Santino CG, Sato SJ, Schuler W, Sims SR, Stehling S, Tarochione LJ, Fromm ME (1995) Field evaluation of european corn borer control in progeny of 173 transgenic corn events expressing an insecticidal protein from *Bacillus thuringiensis*. Crop Sci 35:550–557

Bakan B, Melcion D, Richard-Molard D, Cahagnier B (2002) Fungal growth and fusarium mycotoxin content in isogenic traditional maize and genetically modified maize grown in France and Spain. J Agric Food Chem 50:728–731

Barry BD, Darrah LL, Huckla DL, Antonio AQ, Smith GS, O'Day MH (2000) Performance of transgenic corn hybrids in Missouri for insect control and yield. J Econ Entomol 93:993–999

Barry G, Kishore G, Padgette S, Taylor M, Kolacz K, Weldon M, Re D, Eichholtz D, Fincher K, Hallas L (1992) Inhibitors of amino acid biosynthesis: strategies for imparting glyphosate tolerance to crop plants. Biosynthesis and molecular recognition of amino acids in plants. 139–145

Battraw M, Hall TC (1991) Stable transformation of Sorghum bicolor protoplasts with chimeric neomycin phosphotransferase-II and ß-glucuronidase genes. Theor Appl Genet 82:161–168

Beachy RN, Fraley RT, Rogers SG (2003) Protection of plants against viral infection. Monsanto Technology LLC/Washington University, United States Patent No. 6,608,421

Belanger FC, Kriz AL (1991) Molecular basis for allelic polymorphism of the maize Globulin-1 gene. Genetics 129:863–872

Brauer EK, Shelp BJ (2010) Nitrogen use efficiency: re-consideration of the bioengineering approach. Botany 88:103–109

Cai H, Zhou Y, Xiao J, Li X, Zhang Q, Lian X (2009) Overexpressed glutamine synthetase gene modifies nitrogen metabolism and abiotic stress responses in rice. Plant Cell Rep 28:527–537

Cai T, Daly B, Butler L (1987) Callus induction and plant regeneration from shoot portions of mature embryos of high tannin sorghum. Plant Cell Tiss Org Cult 9:245–252

Cai T, Pierce DA, Tagliani LA, Zhao Z-Y (2002) *Agrobacterium* mediated transformation of sorghum. Pioneer Hi-Bred International, Inc, United States Patent No. 6,396,298

Casas AM, Kononowicz AK, Zehr UB, Tomes DT, Axtell JD, Butler LG, Bressan RA, Hasegawa PM (1993) Transgenic sorghum plants via microprojectile bombardment. Proc Natl Acad Sci U S A 90:11212–11216

Cattaneo MG, Yafuso C, Schmidt CW, Huang C-Y, Rahman M, Olson C, Ellers-Kirk C, Orr BJ, Matsh SE, Antilla L, Dutilleul P, Carriére Y (2006) Farm-scale evaluation of the impacts of transgenic cotton on biodiversity, pesticide use, and yield. Proc Natl Acad Sci U S A 103: 7571–7576

Century K, Reuber TL, Ratcliffe OJ (2008) Regulating the regulators: the future prospects for transcription-factor-based agricultural biotechnology products. Plant Physiol 147:20–29

Christensen AH, Sharrock RA, Quail PH (1992) Maize polyubiquitin genes: structure, thermal perturbation of expression and transcript splicing, and promoter activity following transfer to protoplasts by electroporation. Plant Mol Biol 18:675–689

Coleman CE, Larkins BA (1999) The prolamins of maize. In: Shewry PR, Casey R (eds) Seed proteins. Kluwer Academic Publishers, Dordrecht, pp 109–139

Coleman CE, Lopes MA, Gillikin JW, Boston RS, Larkins BA (1995) A defective signal peptide in the maize high-lysine mutant floury 2. Proc Natl Acad Sci U S A 92:6828–6831

Cook D, Rimando AM, Clemente TE, Schroder J, Dayan FE, Nanayakkara D, Pan Z, Noonan BP, Fishbein M, Abe I, Duke SO, Baerson SR (2010) Alkylresorcinol. synthases expressed in *Sorghum bicolor* root hairs play an essential role in the biosynthesis of the allelopathic benzoquinone soroleone. Plant Cell 22:867–887

Dai S, Zheng P, Marmey P, Zhang S, Tian W, Chen S, Beachy RN, Fauquet C (2001) Comparative analysis of transgenic rice plants obtained by *Agrobacterium*-mediated transformation and particle bombardment. Mol Breed 7:25–33

Daley M, Knauf VC, Summerfelt KR, Turner JC (1998) Co-transformation with one *Agrobacterium tumefaciens* strain containing two binary plasmids as a method for producing marker-free transgenic plants. Plant Cell Rep 17:489–496

Duodu KG, Taylor JRN, Belton PS, Hamaker BR (2003) Factors affecting sorghum protein digestibility. J Cereal Sci 38:117–131

Ejeta G (2007) Breeding for striga resistance in sorghum: exploitation of an intricate host-parasite biology. Crop Sci 47(suppl 3):216–227

Elkonin LA, Pakhomova NV (2000) Influence of nitrogen and phosphorus on induction embryogenic callus of sorghum. Plant Cell Tiss Org Cult 61:115–123

Fraley RT, Rogers SG, Horsch RB, Sanders PR, Flick JS, Adams SP, Bittner ML, Brand LA, Fink CL, Fry JS, Galluppi GR, Goldberg SB, Hoffmann NL, Woo SC (1983) Expression of bacterial genes in plant cells. Proc Natl Acad Sci U S A 80:4803–4807

Gao C, Long D, Lenk I, Nielsen KK (2008) Comparative analysis of transgenic tall fescue (*Festuca arundinacea* Schreb.) plants obtained by *Agrobacterium*-mediated transformation and particle bombardment. Plant Cell Rep 27:1601–1609

Gao Z, Jayaraj J, Muthukrishnan S, Claflin L, Liang GH (2005a) Efficient genetic transformation of sorghum using a visual screening marker. Genome 48:321–333

Gao Z, Xie X, Ling Y, Muthukrishnan S, Liang GH (2005b) *Agrobacterium* tumefaciens-mediated sorghum transformation using a mannose selection system. Plant Biotechnol J 3:591–599

Ghali R, Ghorbel H, Hedilli A (2009) Fumonisin determination in Tunisian foods and feeds. ELISA and HPLC methods and comparison. J Agric Food Chem 57:3955–3960

Gilbert RA, Gallo-Meagher M, Comstock JC, Miller JD, Jain M, Abouzid A (2005) Agronomic evaluation of sugarcane lines transformed for resistance to *sugarcane mosic virus* strain E. Crop Sci 45:2060–2067

Good AG, Johnson SJ, De Pauw M, Carroll RT, Savidov N, Vidmar J, Lu Z, Taylor GJ, Stroeher V (2007) Engineering nitrogen use efficiency with alanine aminotransferase. Can J Bot 85: 252–262

Good AG, Shrawat AK, Muench DG (2004) Can less yield more? Is reducing nutrient input into the environment compatible with maintaining crop production? Trends Plant Sci 9:597–605

Gritz L, Davies J (1983) Plasmid-encoded hygromycin B resistance: the sequence of hygromycin B phosphotransferase gene and its expression in *Escherichia coli* and *Saccaromyces cerevisiae*. Gene 26:179–188

Gurel S, Gurel E, Kaur R, Wong J, Meng L, Tan H-Q, Lemaux PG (2009) Efficient, reproducible *Agrobacterium*-mediated transformation of sorghum using heat treatment of immature embryos. Plant Cell Rep 28:429–444

Hammond BG, Campbell KW, Pilcher CD, Degooyer TA, Robinson AE, McMillen BL, Spangler SM, Riordan SG, Rice LG, Richard JL (2004) Lower fumonisin mycotoxin levels in the grain of *Bt* corn grown in the United States in 2000–2002. J Agric Food Chem 52:1390–1397

Hartings H, Maddaloni M, Lazzaroni N, Di Fonzo N, Motto M, Salamini F, Thompson R (1989) The *O2* gene which regulates zein deposition in maize endosperm encodes a protein with structural homologies to transcriptional activators. EMBO J 8:2795–2801

Hennigh DS, Al-Khatib K, Currie RS, Tuinstra MR, Geier PW, Stahlman PW, Claassen MM (2010) Weed control with selected herbicides in acetolactate synthase-resistant sorghum. Crop Protect 29:879–883

Henzell RG, Persley DM, Greber RS, Fletcher DS, Van Slobbe L (1982) Development of grain sorghum lines with resistance to sugarcane mosaic and other sorghum diseases. Plant Dis 66:900–901

Hokanson KE, Ellstrand NC, Ouedraogo JT, Olweny PA, Schaal PA, Raybould AF (2010) Biofortified sorghum in Africa: using problem formulation to inform risk assessment. Nat Biotechnol 28:900–903

Howe A, Sato S, Dweikat I, Fromm M, Clemente T (2006) Rapid and reproducible *Agrobacterium*-mediated transformation of sorghum. Plant Cell Rep 25:784–791

Huang S, Adams WR, Zhou Q, Malloy KP, Voyles DA, Anthony J, Kriz AL, Luethy MH (2004) Improving nutritional quality of maize proteins by expressing sense and antisense zein genes. J Agric Food Chem 52:1958–1964

Jacob SS, Veluthambi K (2002) Generation of selection marker-free transgenic plants by cotransformation of a cointegrate vector T-DNA and a binary vector T-DNA in one *Agrobacterium tumefaciens* strain. Plant Sci 163:801–806

Jensen SG, Giorda LM (2002) Virus diseases of sorghum and millet in the Americas and Australia. In: Leslie JF (ed) Sorghum and millets diseases. Iowa State, Ames, IA, pp 403–410

Jeoung JM, Krishnaveni S, Muthukrishnan S, Trick HN, Liang GH (2002) Optimization of sorghum transformation parameters using genes for green fluorescent protein and ß-glucuronidase as visual markers. Hereditas 137:20–28

Joersbo M, Okkels FT (1996) A novel principle for selection of transgenic plant cells: positive selection. Plant Cell Rep 16:219–221

Jogeswar G, Ranadheer D, Anjaniah V, Kishor PBK (2007) High frequency somatic embryogenesis and regeneration in different genotypes of *Sorghum bicolor* (L.) Moench from immature explants. In Vitro Cell Dev Biol-Plant 43:159–166

Kaeppler HF, Pedersen JF (1996) Media effects on phenotype of callus cultures initiated from photoperiod-insensitive, elite inbred sorghum lines. Maydica 41:83–89

Karaba A, Dixit S, Greco R, Aharoni A, Trijatmiko KR, Marsch-Martinez N, Nataraja K, Udayakumar M, Pereira A (2007) Improvement of water use efficiency in rice by expression of HARDY, an Arabidopsis drought and salt gene. Proc Natl Acad Sci U S A 104:15270–15275

Khanna H, Becker D, Kleidon J, Dale J (2004) Centrifugation assisted *Agrobacterium* tumefaciens-mediated transformation (CAAT) of embryogenic cell suspensions of banana (*Musa* spp. Cavendish AAA and Lady finger AAB). Mol Breed 14:239–252

Khanna HK, Paul J-Y, Harding RM, Dickman MB, Dale JL (2007) Inhibition of *Agrobacterium*-induced cell death by antiapoptotic gene expression leads to very high transformation efficiency of banana. Mol Plant Microb Interact 20:1048–1054

Komari T, Hiei Y, Saito Y, Murai N, Kumashiro T (1996) Vectors carrying two separate T-DNAs for co-transformation of higher plants mediated by *Agrobacterium tumefaciens* and segregation of transformants free from selection markers. Plant J 10:165–174

Kumar S, Pandey KC (2008) *Bacillus thuringiensis* (Bt) transgenic crop: an environment friendly insect-pest management strategy. J Environ Biol 29:641–653

Maier-Greiner UH, Obermaier-Skrobranek BM, Estermaier LM, Kammerloher W, Freund C, Wulfing C, Burkert UI, Matern DH, Breuer M, Eulitz M et al (1991) Isolation and properties of a nitrile hydratase from the soil fungus *Myrothecium verrucaria* that is highly specific for the fertilizer cyanamide and cloning of its gene. Proc Natl Acad Sci U S A 88:4260–4264

Murashige T, Skoog F (1962) A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol Plant 15:473–497

Negrotto D, Jolley M, Beer S, Wenck AR, Hansen G (2000) The use of phosphomannose-isomerase as a selectable marker to recover transgenic maize plants (*Zea mays* L.) via *Agrobacterium* transformation. Plant Cell Rep 19:798–803

Nelson DE, Repetti PP, Adams TR, Creelman RA, Wu J, Warner DC, Anstrom DC, Bensen RJ, Castiglioni PP, Donnarummo MG, Hinchey BS, Kumimoto RW, Maszle DR, Canales RD, Krolikowski KA, Dotson SB, Gutterson N, Ratcliffe OJ, Heard JE (2007) Plant nuclear factor Y (NF-Y) B subunits confer drought tolerance and lead to improved corn yields on water-limited acres. Proc Natl Acad Sci U S A 104:16450–16455

Nelson RS, Abel PP, Beachy RN (1987) Lesions and virus accumulation in inoculated transgenic tobaco plants expressing the coat protein gene of tobacco mosaic virus. Virology 158:126–132

Nguyen T-V, Thu TT, Claeys M, Angenon G (2007) *Agrobacterium*-mediated transformation of sorghum (*Sorghum bicolor* (L.) Moench) using an improved in vitro regeneration system. Plant Cell Tiss Org Cult 91:155–164

Oria MP, Hamaker BR, Axtell JD, Huang C-P (2000) A highly digestible sorghum mutant cultivar exhibits a unique folded structure of endosperm protein bodies. Proc Natl Acad Sci U S A 97:5065–5070

Palanichelvam K, Oger P, Clough SJ, Cha C, Bent AF, Farrand SK (2000) A second T-region of the soybean-supervirulent chrysopine-type Ti plasmid pTiChry5, and construction of a fully disarmed *vir* helper plasmid. Mol Plant Microb Interact 13:1081–1091

Paterson AH (2008) Genomics of sorghum. Int J Plant Genomics 2008:6

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapmann J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Ginggle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Rahman M, Ware D, Westhoff P, Mayer KFX, Messing J, Rokhsar DS (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556

Pedersen JF, Marx DB, Funnell DL (2003) Use of A3 cytoplasm to reduce risk of gene flow through sorghum pollen. Crop Sci 43:1506–1509

Pola S, Mani NS, Ramana T (2008) Plant tissue culture studies in *Sorghum bicolor*: immature embryo explants as the source material. Int J Plant Prod 2:1–14

Pola SR, Mani NS (2006) Somatic embryogenesis and plantlet regeneration in *Sorghum bicolor* (L.) Moench, from leaf explants. J Cell Mol Biol 5:99–107

Prins M (2003) Broad virus resistance in transgenic plants. Trends Biotechnol 21:373–375

Reddy BN, Raghavender C (2008) Outbreaks of Fusarial-toxicoses in India. Cereal Res Commun 36(Suppl B):321–325

Reddy KRN, Raghavender CR, Reddy BN, Salleh B (2010) Biological control of *Aspergillus flavus* growth and subsequent aflatoxin B_1 production in sorghum grains. Afr J Biotechnol 9:4247–4250

Sahoo L, Schmidt JJ, Pedersen JF, Lee DJ, Lindquist JL (2010) Growth and fitness components of wild X cultivated *Sorghum bicolor* (Poacheae) hybrids in Nebraska. Am J Bot 97:1610–1617

Sato S, Clemente T, Dweikat I (2004a) Identification of an elite sorghum genotype with high in vitro performance capacity. In Vitro Cell Dev Biol 40:57–60

Sato S, Xing A, Ye X, Schweiger B, Kinney A, Graef G, Clemente T (2004b) Production of γ-linolenic acid and stearidonic acid in seeds of marker-free transgenic soybean. Crop Sci 44:646–652

Sharma HC, Ortiz R (2000) Transgenics, pest management, and the environment. Curr Sci 79:421–437

Shrawat AK, Carroll RT, DePauw M, Taylor GJ, Good AG (2008) Genetic engineering of improved nitrogen use efficiency in rice by the tissue-specific expression of *alanine aminotransferase*. Plant Biotechnol J 6:722–732

Stark DM, Beachy RN (1989) Protection against potyvirus infection in transgenic plants: evidence for broad spectrum resistance. Nat Biotechnol 7:1257–1262

Suzuki N, Rizhsky L, Liang H, Shuman J, Shulaev V, Mittler R (2005) Enhanced tolerance to environmental stress in transgenic plants expressing the transcriptional coactivator multiprotein bridging factor $1c^{1[w]}$. Plant Physiol 139:1313–1322

Tesso T, Ejeta G, Chandrashekar A, Huang C-P, Tandjung A, Lewamy M, Axtell JD, Hamaker BR (2006) A novel modified endosperm texture in a mutant high-protein digestibility/high-lysine grain sorghum (*Sorghum bicolor* (L.) Moench). Cereal Chem 83:194–201

Tesso T, Hamaker BR, Ejeta G (2008a) Sorghum protein digestibility is affected by dosage of mutant alleles in endosperm cells. Plant Breed 127:579–586

Tesso T, Kapran I, Grenier C, Snow AA, Sweeney P, Pedersen JF, Marx DB, Bothma G, Ejeta G (2008b) The potential for crop-to-wild gene flow in sorghum in Ethiopia and Niger: a geographic survey. Crop Sci 48:1425–1431

Thompson CJ, Movva NR, Tizard R, Crameri R, Davies JE, Lauwereys M, Botterman J (1987) Characterization of the herbicide-resistance gene *bar* from *Streptomyces hygroscopicus*. EMBO 6:2519–2523

Torres MJ, Tomilov AA, Tomilova N, Reagan RL, Yoder JI (2005) Psroph, a parasitic plant EST database enriched for parasite associated transcripts. BMC Plant Biol 5:24–33

Tuinstra MR, Soumana S, Al-Khatib K, Kapran I, Toure A, van Ast A, Bastiaans L, Ochanda NW, Salami I, Kayentao M, Dembele S (2009) Efficacy of herbicide seed treatments for controlling Striga infestation of sorghum. Crop Sci 49:923–929

Wei H, Moore PH, Albert HH (2003) Comparative expression analysis of two sugarcane polyubiquitin promoters and flanking sequences in transgenic plants. J Plant Physiol 160:1241–1251

Wu Y, Messing J (2010) RNA interference-mediated change in protein body morphology and seed opacity through loss of different zein proteins. Plant Physiol 153:337–347

Xin Z, Wang M-L, Burow G, Burke J (2009) An induced sorghum mutant population suitable for bioenergy research. Bioenerg Res 2:10–16

Xing A, Zhang Z, Sato S, Staswick P, Clemente T (2000) The use of the two T-DNA binary system to derive marker-free transgenic soybeans. In Vitro Cell Dev Biol 36:456–463

Yanagisawa S, Akiyama A, Kisaka H, Uchimiya H, Miwa T (2004) Metabolic engineering with Dof1 transcription factor in plants: Improved nitrogen assimilation and growth under low-nitrogen conditions. Proc Natl Acad Sci U S A 101:7833–7838

Yoder JI, Scholes JD (2010) Host plant resistance to parasitic weeds; recent progress and bottlenecks. Curr Opin Plant Biol 13:478–484

Zhao Z-Y, Glassman K, Sewalt V, Wang N, Miller M, Chang S, Thompson T, Catron S, Wu E, Bidney D, Kedebe Y, Jung R (2003) Nutritionally improved transgenic sorghum. Kluwer Academic Publishers, Dordrecht

Zhao Z-Y, Gu W, Cai T, Tagliani LA, Hondred D, Bond D, Krell S, Rudert ML, Bruce WB, Pierce DA (1998) Molecular analysis of T0 plants transformed by Agrobacterium and comparison of Agrobacterium-mediated transformation with bombardment transformation in maize. Maize Genet Coop News Lett 72:1–4

Zhao ZY, Cai T, Tagliani L, Miller M, Wang N, Pang H, Rudert M, Schroeder S, Hondred D, Seltzer J, Pierce D (2000) *Agrobacterium*-mediated sorghum transformation. Plant Mol Biol 44:789–798

Chapter 11
Genetic Engineering of *Saccharum*

Getu Beyene, Ian S. Curtis, Mona B. Damaj, Marco T. Buenrostro-Nava, and T. Erik Mirkov

Abstract Over the last two decades, substantial progress has been made in the genetic engineering of sugarcane (*Saccharum* spp.) through improvements in tissue culture procedures, allowing a higher efficiency of generating transgenic plants using *Agrobacterium*-mediated and biolistic gene transfers. Elucidation of gene function and development of varieties with improved yield, sugar level, fiber content, and other desirable traits and products for superior performance have been possible through transgenic technologies. Researchers are now focusing on optimizing existing methodologies and developing new technologies for the production of elite varieties, enhancement of transgene expression, and manipulation of metabolic pathways for improved molecular breeding and commercial exploitation. At present, no transgenic sugarcane has been released to the commercial market, but with the aid of large investments from the private sector, the commercialization of this major sugar- and biomass-producing crop should be accelerated.

Keywords *Saccharum* spp. • Biotechnology • Transformation • Improved agronomic traits • Metabolic engineering • Biofactory • Transgene silencing

G. Beyene (✉)
Texas AgriLife Research, Texas A&M System,
College Station, Weslaco, TX 78596, USA

International Laboratory for Tropical Agricultural Biotechnology (ILTAB), Donald Danforth Plant Science Center, 975 North Warson Road, St. Louis, MO 63132, USA
e-mail: gduguma@danforthcenter.org

I.S. Curtis • M.B. Damaj • M.T. Buenrostro-Nava • T.E. Mirkov
Texas AgriLife Research, Texas A&M System,
2415 E. Hwy 83, Weslaco, TX 78596, USA

A.H. Paterson (ed.), *Genomics of the Saccharinae*, Plant Genetics and Genomics: Crops and Models 11, DOI 10.1007/978-1-4419-5947-8_11,

1 Historical Perspective

Sugarcane (*Saccharum* spp.) has a complex aneuploid genome with a large and variable number of chromosomes per cell (D'Hont et al. 1998). Such a complex genome results in the introgression and selection of desirable traits into elite clones through conventional breeding to be a long-term undertaking (10–15 years). Genetic engineering offers the potential of introducing traits into sugarcane in a timely manner. The early establishment of a reproducible in vitro cell and tissue culture system for sugarcane has facilitated the development of efficient methods for its genetic transformation. Tissue culture has emerged as a valuable tool since the first report on plant regeneration from induced callus (Barba and Nickell 1969; Heinz and Mee 1969). Tissue culture protocols were later refined and incorporated into different programs aiming at cell improvement through mutation (Patade and Suprasana 2008; Patade et al. 2008), micropropagation (Behera and Sahoo 2009; Khan et al. 2008; Lee 1987; Roy and Kabir 2007), production of disease-free plants (Fitch et al. 2001; Irvine and Benda 1985), and genetic engineering.

Early studies on sugarcane transformation involved DNA transfer into protoplasts as target cells by polyethylene glycol (PEG) treatment (Chen et al. 1987) or via electroporation (Chowdhury and Vasil 1992; Rathus and Birch 1992) to produce the first transgenic sugarcane cell lines. However, no transgenic plants could be recovered with either technique, as regeneration is very tedious from electroporated protoplasts, and inefficient and poorly reproducible with PEG-transformed protoplasts. The first transgenic sugarcane plants were obtained following electroporation of meristematic tissues of in vitro grown plants (Arencibia et al. 1992) and then later intact embryogenic callus (Arencibia et al. 1995, 1997). Simultaneously, biolistic gene transfer (developed in the late 1980s; reviewed by Altpeter et al. 2005) was also successfully adopted for the generation of the first transgenic plants from cell suspensions and embryogenic calli (Bower and Birch 1992; Irvine and de Almedia 1991), and a refined protocol for the transformation of a commercial sugarcane variety was described (Bower and Birch 1992). This marked a major milestone in sugarcane transformation. Further refinements in tissue culture media and explant preparation and improvements in biolistic transfer protocols allowed transformation of sugarcane to become fairly routine in many laboratories (Tables 11.1 and 11.2), using easily established target tissues, including embryogenic callus (Bower et al. 1996; Falco et al. 2000; Gallo-Meagher and Irvine 1996; Gilbert et al. 2005; Ingelbrecht et al. 1999; Joyce et al. 1998; Snyman et al. 1996; Weng et al. 2006; Zhang et al. 1999) and immature leaf roll disk explants (Snyman et al. 2000, 2006). *Agrobacterium*-mediated gene transfer was successful in sugarcane for a limited number of varieties and target tissues such as embryogenic calli (Arencibia et al. 1998; Elliott et al. 1998), immature leaf sections (Enríquez-Obregón et al. 1997, 1998), and axillary meristems (Manickavasagam et al. 2004).

The first efforts on sugarcane improvement using genetic engineering concentrated on the production of herbicide and biotic stress resistant transgenic plants (Table 11.1).

Table 11.1 Transgenic sugarcane with input traits

Herbicide resistance

Glufosinate ammonium

- *bar*, Maize *Ubi-1*/*nos*:
  - Biolistic, Bialaphos (1–3 mg/l), Field trial, NCo 310 (Gallo-Meagher and Irvine 1996)
  - *Agrobacterium*, Phosphinothricin (4 mg/l), Field trial, Ja60-5 (Enríquez-Obregón et al. 1998)
  - Biolistic, Geneticin (30 mg/l), No field trial, SP80-180 (Falco et al. 2000)
- *bar*, *CaMV 35S*, *Agrobacterium*, Phosphinothricin (5 mg/l), No field trial, Co 92061 and Co 671 (Manickavasagam et al. 2004)
- *pat*, Maize *Ubi-1*, Biolistic, Glufosinate ammonium, No field trial, NCo 310 (Leibbrandt and Snyman 2003)

Biotic stress tolerance

Insect

- Sugarcane stem borer:
  - *cry1A(b)*, *CaMV 35S*/*nos*, Electroporation, GUS, Field trial, Ja60-5 (Arencibia et al. 1997, 1999)
  - *cry1A(b)*, maize phopshoenolpyruvate carboxylase promoter and pith promoter, Biolistic, Neomycin, Field trial, SP80-1842 and SP80-3280 (Braga et al. 2003)
  - *cry1A(c)*, Maize *Ubi-1*/*nos*, Biolistic, Geneticin (50 mg/l), No field trial, YT79-177 and ROC16 (Weng et al. 2006)
- Sugarcane canegrub, *gna* or *pinII*, Maize *Ubi-1*/*nos*, Biolistic, Geneticin, No field trial, Q117 (Nutt et al. 1999)
- Mexican rice borer, *gna*, Maize *Ubi-1*/*nos*, Biolistic, Geneticin (30 mg/l), Field trial, CP65-357 (Legaspi and Mirkov 2000)
- Sugarcane stem borer and Mexican rice borer, *gna*, Maize *Ubi-1*/*nos*, Biolistic, Geneticin (30 mg/l), Field trial, CP65-357 (Sétamou et al. 2002)
- Sugarcane borer, *SKTI* and *SBBI*, Maize *Ubi-1*/*nos*, Biolistic, Geneticin (30 mg/l), No field trial, SP80-1842 and SP80-3280 (Falco and Silva-Filho 2003)
- Sugarcane top borer, *aprotinin*, Maize *Ubi-1*/*nos*, Biolistic, Hygromycin (30–50 mg/l), No field trial, CoC 92061 and Co 86032 (Christy et al. 2009)
  - Sugarcane shoot borer, *cryIA(b)* or *aprotinin, Maize Ubi-1*/nos, Biolistic or *Agrobacterium*, Phosphinothricin (3 mg/l), No field trial, Co 86032 and CoJ 64 (Arvinth et al. 2010)

Disease

- Sugarcane mosaic viral (SCMV) disease:
  - *SCMV-CP*, Synthetic *Emu*/*nos* or Maize *Ubi-1*/*nos*, Biolistic, Geneticin (60 mg/l), No field trial, Q155, Q95 and Q153 (Joyce et al. 1998)
  - *SCMV-CP*, Maize *Ubi-1*/*nos*, Biolistic, Geneticin (15–45 mg/l), Field trial, CP84-1198 (Gilbert et al. 2005)
- Sorghum mosaic viral (SrMV) disease, *SrMV-CP*, Maize *Ubi-1*/*nos*, Biolistic, Geneticin (15–45 mg/l) and Bialaphos (1–3 mg/l), No field trial, CP72-1210 and CP65-357 (Ingelbrecht et al. 1999)
- Sugarcane leaf scald, *albD*, Maize *Ubi-1*/*nos*, Biolistic, Geneticin, Field trial, Q63 and Q87 (Zhang et al. 1999)
- Fiji leaf gall, *FDVS9 ORF 1*, Maize *Ubi-1*/*nos*, Biolistic, Geneticin (60 mg/l), No field trial, Q124 (McQualter et al. 2004)
- Sugarcane yellow leaf viral (SCYLV) disease, *SCMV-CP*, Maize *Ubi-1*/*nos*, Biolistic, Geneticin (50–60 mg/l), Field trial, CP92-1666 (Gilbert et al. 2009)

(continued)

Table 11.1 (continued)

Abiotic stress tolerance
Drought
• Trehalose production, *TSase*, 2*XCaMV 35S/nos*, *Agrobacterium*, Phosphinotricin (0.5–0.75 mg/l), Field trial, ROC10 (Zhang et al. 2006)
• Proline production, *Vap5cs*, ABA-inducible promoter complex, Biolistic, Glufosinate ammonium (5 mg/l), No field trial, RB855156 (Molinari et al. 2007)
• Drought regulation, *AtCBF4*, Maize *Ubi-1/nos*, Biolistic, Geneticin (50 mg/l), No field trial, Q117 (McQualter and Dookun-Saumtally 2007)

Information on each trait is presented as follows: trait, gene, promoter/terminator, transformation method, selection agent, field trial, variety and citation. *bar*, bialaphos resistance gene; *Ubi-1*, *ubiquitin-1*; *nos*, *nopaline synthase* terminator; *CaMV 35S*, *Cauliflower mosaic virus 35S*; *pat*, *phosphinothricin acetyltransferase* gene; *cry*, crystal protein gene; GUS, β-glucuronidase; *gna*, *Galanthus nivalis agglutinin*; *pinII*, potato *proteinase inhibitor II*; *SKTI*, soybean Kunitz *trypsin inhibitor*; *SBBI*, soybean Bowman-Birk *trypsin inhibitor*; *Emu*, a truncated maize *alcohol dehydrogenase* (*Adh1*) promoter with copies of the anaerobic responsive element of the maize *Adh1* gene and ocs-elements of the *Agrobacterium octopine synthase* gene; *albD*, albicidin detoxifying gene; *FDVS9 ORF 1*, *Fiji disease virus* segment 9 ORF 1; *TSase*, *trehalose synthase*; *AtCBF4*, *Arabidopsis thaliana* C-repeat binding factor 4; *Vap5cs*, *Vigna aconitifolia* Δ^1*-pyrroline-5-carboxylate synthetase*

Metabolic engineering for other economically important traits such as enhancing sucrose content and using transgenic sugarcane to produce novel and value-added products represent a more recent era of genetic improvement (Table 11.2).

2 Current Methods

Genetic engineering offers a practical approach of introducing specific traits into the genome of desired varieties. To date, the transfer of foreign DNA into sugarcane has been achieved by PEG-treatment, electroporation, *Agrobacterium*-mediated transformation and particle bombardment. Here, we describe the latter two procedures due to their popularity.

2.1 Agrobacterium-Mediated Transformation

In general, the use of *Agrobacterium* as the vehicle for transforming plant cells is the favored method over biolistics in terms of more stable transgene expression and the integration of fewer copies of the transgene, which might help to reduce the incidence of transgene silencing as seen in studies on rice (Dai et al. 2001). Despite the advantages of using *Agrobacterium* for producing transgenic plants, conditions that affect the generation of transformed sugarcane plants have been poorly investigated compared to other monocotyledonous plants. Below, we outline the factors that have been studied in the optimization of producing transformed sugarcane events using *Agrobacterium*.

To date, embryogenic callus represents the major source of sugarcane material used for *Agrobacterium*-inoculation in the production of transgenic plants, and it is

Table 11.2 Transgenic sugarcane with output traits

Metabolic engineering

Sugar metabolism

- Sucrose:
  - *Soluble acid invertase*, Maize *Ubi-1*/*nos*, Biolistic (cell suspension), Geneticin (50–100 mg/l), No field trial, H62-4671 (Ma et al. 2000)
  - *Neutral invertase* (antisense), *CaMV 35S*- maize *Ubi-1*, Biolistic, No field trial, NCo 310 (Rossouw et al. 2007)
  - *Soluble acid invertase*, *CaMV 35S*-Maize *Ubi-1*, Biolistic, No field trial, Q117 (Botha et al. 2001)
  - *PFP*, *CaMV 35S*-Maize *Ubi-1*/*nos*, Biolistic, Geneticin, No field trial, NCo 310 (Groenewald and Botha 2008)

Secondary metabolism

- Hydroxybenzoic acid, *hchl* and *cpl*, Maize *Ubi-1*/*nos*, Biolistic, Geneticin (60 mg/l), No field trial, Q117 (McQualter et al. 2005)
- Polyphenol oxidase, *ppo*, Maize *Ubi-1*/*nos*, Biolistic, Geneticin, Field trial, Q117 (Vickers et al. 2005a)

Production of value-added products

Sugar by-products

- Sorbitol, *mds6pdh*, Maize *Ubi-1*/*nos*, Biolistic, Paramomycin (150 mg/l), No field trial, Q117 (Chong et al. 2007)
- Isomaltulose, *UQ68J SI*, Maize *Ubi-1*/*nos*, Biolistic, Geneticin, No field trial, Q117 (Wu and Birch 2007)

Other products

- Polyhydroxybutyrate (PHB) biopolymer, *phaA*, *phaB* and *phaC*, Maize *Ubi-1*/*nos*, Biolistic, Geneticin, No field trial, Q117 (Petrasovits et al. 2007)
- Human cytokine granulocyte macrophage colony stimulating factor (GM-CSF), Maize *Ubi-1*/*nos*, Biolistic, Geneticin (70–100 mg/l), Field trial, H62-4671 and Q117 (Wang et al. 2005a)

Information on each trait is presented as follows: trait, gene, promoter/terminator, method of transformation, selection agent, field trial, variety and citation. *Ubi-1*, *ubiquitin-1*; *nos*, *nopaline synthase* terminator; *CaMV 35S*, *Cauliflower mosaic virus 35S*; *PFP*, *fructose 6-phosphate 1-phosphotransferase*; *hchl*, *Pseudomonas fluorescens 4-hydroxycinnamoyl-CoA hydratase/lyase*; *cpl*, *Escherichia coli chorismate pyruvate-lyase*; *ppo*, *polyphenol oxidase*; *mds6pdh*, *Malus domestica sorbitol-6-phosphate dehydrogenase*; *UQ68J SI*, *sucrose isomerase* from *Pantoea dispersa UQ68J*; *phaA*, *Ralstonia eutropha β-ketothiolase*; *phaB*, *R. eutropha acetoacetyl-reductase*; *phaC*, *R. eutropha polyhydroxybutyrate synthase*

achieved by culturing leaf roll sections or immature inflorescences on an auxin-containing medium, normally 2,4-dichlorophenoxyacetic acid (2,4-D), at a concentration of 3 mg/l. However, a recent report has shown that plants can be initiated through the inoculation of axillary bud explants via direct regeneration in Indian varieties of sugarcane, allowing phenotypically normal progenies to be transformed (Manickavasagam et al. 2004). The type and age of the callus were found to be critical in determining transformation efficiency, with young (approx. 4 weeks) regenerable calli being more receptive to gene transfer with *Agrobacterium* (Arencibia et al. 1998). In a separate study, preconditioning of meristematic explants of sugarcane with antinecrotics to minimize oxidative burst improved transformation efficiencies markedly (Enríquez-Obregón et al. 1998). Nevertheless,

it is important to note that such studies on explant type and conditioning required other intervening steps for the optimization of transgenic plant generation, which further complicates the development of the gene transfer system.

The use of a selectable marker in selecting transformation events is essential when designing a gene transfer system in sugarcane. A report using green fluorescent protein (*gfp*) gene to follow *Agrobacterium*-mediated transformation in a selectable marker-free method resulted in very few transformation events (Elliott et al. 1998). The low transformation efficiency was due to the lack of selective growth advantage of transformed over nontransformed events. To date, the markers used in selecting for transformed plants include the hygromycin phosphotransferase gene (*hph*) or the bialaphos resistance *bar* gene (Table 11.1). However, more recently, the neomycin phosphotransferase II (*npt*II) gene has been successfully shown to be efficient in selecting transformed sugarcane callus (Zhangsun et al. 2007). Transformed shoots of the commercial variety Q117 were only regenerated when calli were inoculated with an *Agrobacterium* strain harboring a plasmid with the *npt*II gene as opposed to the one with the *hph* gene, suggesting that the selection system is important for producing transgenic shoots (Joyce et al. 2010). The advantage of using the *npt*II gene in transformation studies is that the NPTII protein is nontoxic to humans and transgenic plants containing this gene have been declared safe for release by the US Environmental Protection Agency (1994).

Several strains of *A. tumefaciens* were successfully used for transformation of sugarcane, including LBA4404, EHA101, EHA105, AGL0, and AGL1 (Arencibia et al. 1998; Elliott et al. 1998; Joyce et al. 2010; Manickavasagam et al. 2004). Cocultivation of embryogenic callus with each of LBA4404, EHA105, AGL0, and AGL1 strains on Murashige and Skoog (MS) (1962) based media for 4 days yielded the highest number of transformed shoots (Joyce et al. 2010). Addition of 2,4-D in cocultivation media was important in the production of transgenic sugarcane using the strains EHA105, AGL0, and AGL1, but not LBA4404. Finally, inoculation conditions that were optimal for routine T-DNA delivery into calli such as the use of vacuum infiltration and wounding calli with particles fired by a biolistic gun were not ideal for tissue survival and plant regeneration in sugarcane (Joyce et al. 2010). Optimization of *Agrobacterium*-mediated transformation of sugarcane is under way in many laboratories that have adopted this technique on a routine basis.

2.2 Particle Bombardment

Particle bombardment involves coating of DNA on small (0.6–1.2 μm) carrier particles of gold or tungsten, which is then propelled at high velocity to a target tissue in a partial vacuum chamber using pressurized helium or electric discharge. The impact force from high velocity propelled microparticles breaks the plant cell wall and plasma membrane and allows the introduction of the DNA into or adjacent to the nucleus. This technique is very simple and applicable to a diverse range of regenerable tissues targeted for transformation (Birch and Franks 1991; Franks and Birch 1991; Lakshmanan et al. 2005). Early studies in the development and optimization of

sugarcane transformation identified the target tissue and the method for selection and regeneration of transformants as the most important factors for the production of stably transformed plants by particle bombardment (Birch and Franks 1991).

2.2.1 Target Tissue

Since the generation of the first transgenic sugarcane plants from particle bombarded embryogenic callus as target tissue (Bower and Birch 1992), protocols for particle bombardment-mediated transformation of sugarcane relied on the use of embryogenic callus (Bower and Birch 1992; Bower et al. 1996; Snyman 2004), even though meristematic tissues (Gambley et al. 1993, 1994) and suspension cells (Chowdhury and Vasil 1992; Franks and Birch 1991) were successfully transformed. More recently, young leaf roll disks with or without preemergent inflorescence (also called immature inflorescence) have emerged as a suitable target for transformation (Snyman et al. 2000, 2006). Young leaf roll disks are common explants used for the initiation and establishment of embryogenic callus or direct somatic embryogenesis. Such leaf roll disks are obtained as transverse sections of the apical portion of 4- to 8-month-old sugarcane tops just above the apical meristem, and are cultured on MS medium supplemented with the auxin 2,4-D (Bower and Birch 1992; Ho and Vasil 1983; Snyman 2004). Embryogenic cells are initiated directly on the explants or indirectly from undifferentiated callus depending on the concentration of 2,4-D used and length of time in culture. At low concentrations of 2,4-D (0.25–0.6 mg/l), somatic embryogenesis is induced directly on leaf roll disks within 2–4 weeks (Desai et al. 2004; Ho and Vasil 1983; Snyman et al. 2000, 2006). At high concentrations of 2,4-D (3 mg/l), embryogenic tissues are produced indirectly from undifferentiated callus within 8–12 weeks (Bower and Birch 1992; Snyman 2004).

The choice of the target tissue (embryogenic callus or embryogenic leaf roll disk) determines not only the time needed to generate a transgenic plant but also the efficiency of the transformation process (defined as the number of transgenic plants generated per bombardment). Snyman et al. (2000) reported an 18-fold increase (0.1 vs. 1.8 plants) in transformation efficiency and a timesaving of 10–14 weeks when embryogenic leaf roll disks were targeted as compared to embryogenic callus. Furthermore, targeting preemergent inflorescences improved transformation efficiency significantly (0.6–5.25 plants) as compared to young leaf roll disks (0–0.2 plants) (Snyman et al. 2006). Improvement in transformation efficiency was attributed to a higher embryogenic competence of inflorescence tissue (Snyman et al. 2006). Besides improving transformation efficiency, direct embryogenesis from leaf roll disks reduces the duration of exposure of explants to high levels of 2,4-D and minimizes the incidence of somaclonal variation (Snyman et al. 2006) observed in transgenic plants generated from embryogenic callus (Gilbert et al. 2005, 2009; Vickers et al. 2005b). However, one of the pitfalls of direct embryogenesis from leaf roll disks is the high incidence of escapes (Snyman et al. 2000, 2006), possibly due to low surface area exposure of the embryogenic leaf roll disks to the selective agent during selection and regeneration.

2.2.2 Selection and Regeneration of Transformants

Selection and regeneration are determining steps for the efficient development of transgenics. At present, plant transformation methods have demonstrated that the success of introducing foreign DNA into the cell is estimated to be 1 in 1,000 to 1 in one million of treated cells (Brasileiro and Aragão 2001). Hence, an efficient selection system is required to recover the low frequency of transformed cell lines. The two selection systems adopted for sugarcane transformation are based on the use of either visible/scorable reporter genes, or selectable markers conferring resistance to antibiotics or herbicides.

Reporter/scorable markers such as β-glucuronidase (GUS), GFP, yellow fluorescent protein (EYFP), luciferase (LUC), and maize anthocyanin (ANT) regulatory elements have been used for the development and optimization of transformation to allow visual selection of transient and stable integrations (Arencibia et al. 1995; Bower and Birch 1992; Bower et al. 1996; Elliott et al. 1998), functional analysis (Braithwaite et al. 2004; Damaj et al. 2010; Gallo-Meagher and Irvine 1993; Liu et al. 2003; Wei et al. 2003), subcellular targeting studies (Gnanasambandam and Birch 2004; Gnanasambandam et al. 2007), dissection of transgene silencing (Birch et al. 2010a), and determination of terminator efficiencies of constructs (Beyene et al. 2011).

The most common selectable marker used in stable sugarcane transformation is the antibiotic resistance *npt*II gene (Tables 11.1 and 11.2), originally isolated from the *Escherichia coli* Tn5 transposon (Beck et al. 1982). This gene confers resistance to neomycin, kanamycin, and a number of aminoglycoside analogues such as paromomycin, geneticin (G418), and butirosin (Goldstein et al. 2005). However, G418 is used as a selection agent since sugarcane has natural resistance to kanamycin (Arencibia et al. 1992; Hauptmann et al. 1988). G418 is applied at a concentration range of 15–100 mg/l as stepwise increase or a single dosage throughout selection and regeneration stages (Tables 11.1 and 11.2). Paromomycin has been occasionally used as a selection agent at a concentration of 150 mg/l (Chong et al. 2007).

The herbicide resistance *bar* (Thompson et al. 1987) and *pat* (Strauch et al. 1988) genes are also common selection markers used in stable transformation of sugarcane. They code for the enzyme phosphinothricin acetyltransferase (PAT) that detoxifies phosphinothricin (PPT), the active ingredient of commercial herbicides such as Basta™ and Liberty™ by acetylation (De Block et al. 1987). Bialaphos (1–3 mg/l) (Gallo-Meagher and Irvine 1996; Ingelbrecht et al. 1999) or glufosinate ammonium (5 mg/l) (Molinari et al. 2007) is used for selecting *bar* transformed sugarcane events (Table 11.1). SeedChek™ Liberty™ Link Test Strips (Strategic Diagnostics Inc. Newark, DE) are also available for early detection of PAT expressing events in tissue culture plants.

Nonantibiotic and nonherbicide selection systems based on enzymes involved in intermediary metabolism are also available, including the *E. coli manA* gene encoding for the enzyme phosphomannose isomerase (PMI). PMI is not present in many

plants and catalyzes the reversible interconversion of mannose 6-phosphate and fructose 6-phosphate (Privalle et al. 2000). This selection system is also called positive selection since cells expressing PMI acquire a growth advantage due to their ability to utilize mannose-6-phosphate, which otherwise is toxic to plants when mannose is supplied as a carbon source (Brumbley et al. 2008; Darbani et al. 2007; Privalle et al. 2000). In sugarcane, Jain et al. (2007) were the first to show the successful application of the PMI/mannose selection system, using 3 g/l of mannose in the presence of sucrose during selection and regeneration, and 1.5–3 g/1 of mannose for rooting.

Regeneration of transgenic plantlets from leaf roll disks and embryogenic calli can involve somatic embryogenesis, secondary proliferation of developing embryoids and/or shoot formation and proliferation via organogenesis (Brumbley et al. 2008; Chen et al. 1988; Ho and Vasil 1983; Lakshmanan et al. 2006; Snyman 2004). Following selection, transgenic sugarcane plants are usually generated in absence of growth regulators (Bower and Birch 1992; Snyman et al. 2006), or in presence of auxins and cytokinins such as 2,4-D (1 mg/l) (Gallo-Meagher and Irvine 1996), kinetin (0.5 mg/l) (Snyman 2004), or kinetin (2 mg/l) and naphthalene acetic acid (5 mg/l) (Ingelbrecht et al. 1999). Rooting of regenerated plants is achieved by the supplementation of indole-3-butyric acid (0.5–4 mg/l).

3 Traits Engineered into Sugarcane

Traits introduced into sugarcane can be categorized as "input" and "output" traits (Brumbley et al. 2008; Willmitzer 1999). Input traits influence crop performance and productivity, and include those conferring biotic and abiotic stress tolerance as well as herbicide resistance. Output traits determine the quality and composition of the end product such as enhanced sucrose content and production of other alternative and high-value products. Both input and output traits engineered into sugarcane are presented in Tables 11.1 and 11.2 and have been reviewed in detail elsewhere (Altpeter and Oraby 2010; Brumbley et al. 2008; D'Hont et al. 2008; Lakshmanan et al. 2005; Ming et al. 2006).

3.1 Input Traits

The first input trait engineered into sugarcane was herbicide resistance, which was achieved by introducing the *bar* or *pat* gene (Table 11.1). The first study on herbicide resistance showed that the generated *bar* transgenic lines (variety NCo 310) displayed a stable transgene expression for three vegetative cycles of propagation, although they contained 3–10 transgene copies and varied phenotypically from very susceptible to highly resistant to the herbicide (Gallo-Meagher and Irvine 1996). Similarly, analysis

of the herbicide resistant *pat* expressing plants from a single transgenic line revealed a stable transgene expression over three generations of vegetative propagation, with no significant differences in measured agronomic characters from the untransformed counterpart under field conditions (Leibbrandt and Snyman 2003). Based on a limited scale field trial data, Enríquez-Obregón et al. (1998) reported a varying level of herbicide resistant plants expressing the *bar* gene, despite the presence of 1–2 transgene copies.

Sugarcane was also engineered for tolerance to different strains of *Sugarcane mosaic virus* (SCMC) (Gilbert et al. 2005; Joyce et al. 1998), *Sorghum mosaic virus* (SrMV) (Ingelbrecht et al. 1999), *Sugarcane yellow leaf virus* (SCYLV) (Gilbert et al. 2009) and *Fiji disease virus* (McQualter et al. 2004) (Table 11.1). Gilbert et al. (2005) detected a large variability in field disease tolerance and agronomic performance among a population of 100 plants from the varieties CP84-1198 and CP80-1827 transformed for tolerance to SCMV strain E. The large number of transformed events allowed for a more rigorous selection for improved agronomic performance and tolerance to the virus over the corresponding controls. By generating two transgenic sugarcane lines for SCYLV tolerance, Gilbert et al. (2009) were able to obtain significantly higher levels of disease tolerance compared to nontransformed controls. However, the agronomic performance of these lines was low.

Insect tolerant transgenic sugarcane engineered with the *Bacillus thuringiensis* (*Bt*) toxin genes *cry1A(b)* and *cry1A(c)* (Arencibia et al. 1997, 1999; Arvinth et al. 2010; Braga et al. 2003; Weng et al. 2006), the carbohydrate binding lectin gene *gna* and the proteinase inhibitor genes *pinII*, *SKTI*, *SBBI*, and *aprotinin* (Christy et al. 2009; Legaspi and Mirkov 2000; Nutt et al. 1999; Sétamou et al. 2002) were also generated (Table 11.1), and field or greenhouse trials have demonstrated encouraging levels of insect tolerance.

Limited reports are available on transgenic sugarcane engineered for abiotic stress tolerance (Table 11.1). In one study, drought-tolerant transgenic sugarcane was generated by the expression of a proline biosynthetic gene (*Vap5cs*) from *Vigna aconitifolia* under the control of an ABA-inducible promoter complex (Molinari et al. 2007). The transgenic plants accumulated similar levels of proline and performed the same as wild type under well-watered conditions; however, under water deficit stress, the transgenic plants accumulated 2.5-fold higher proline levels, and exhibited a 65 % increase in photochemical efficiency of PSII and a higher biomass production than nontransformed plants. In a similar study, overexpression of the *Grifola frondosa* trehalose synthase (*TSase*) gene under the control of two tandem copies of the *Cauliflower mosaic virus* (*CaMV*) *35S* promoter in sugarcane resulted in a high accumulation of trehalose (8.8–12.8 mg/g fresh weight) and an increased drought tolerance under field conditions, with no detectable alteration to plant growth and development (Zhang et al. 2006).

3.2 Output Traits

Sugarcane is unique among other plants in its ability to store sucrose, the commercial product, at unprecedented levels (up to 42 % dry weight of stem) (Brumbley et al. 2008).

Besides its ability to produce a high biomass, sugarcane possesses inbuilt genetic containment features. It is vegetatively propagated and does not produce seeds or pollen under some commercial settings and as such pollen-mediated vertical and horizontal gene transfers are unlikely to occur. Early extraction procedures to obtain sucrose from sugarcane plants remove any contaminating nucleic acids and proteins avoiding biosafety concerns of genetically modified sugarcane. All of these features make this crop an attractive choice for biofactory applications (Altpeter and Oraby 2010; Birch 2007; Brumbley et al. 2008; D'Hont et al. 2008; Ming et al. 2006).

Sugarcane accounts for over 70 % of the world's sugar demand and hence the improvement in sucrose content has been one of the primary targets for researchers. A number of genes involved in sucrose metabolism have been identified and targeted for enhanced sucrose content (see reviews: Altpeter and Oraby 2010; Brumbley et al. 2008; D'Hont et al. 2008; Lakshmanan et al. 2005; Ming et al. 2006) (Table 11.2). However, attempts to alter sucrose content by overexpressing targeted genes had mixed success due to the complex interactions among the different processes involved (Moore 2005). One of the most important achievements to be noted is the production of a sweetener, isomaltulose. A gene encoding a bacterial enzyme, sucrose isomerase, that converts sucrose to its isomer isomaltulose, was constitutively expressed in the storage vacuoles of sugarcane (Wu and Birch 2007). Isomaltulose accumulated in storage tissues without any decrease in stored sucrose concentrations, resulting in double concentrations of total sugar (isomaltulose and sucrose) in the harvested juice.

Replacement of synthetic plastic with biodegradable equivalents is one of the environmentally benign approaches that biotechnology can offer. The bacterial polyester, polyhydroxybutyrate (PHB) was produced in sugarcane by expressing the PHB biosynthetic enzymes of *Ralstonia eutropha* genes, *phaA*, *phaB*, and *phaC*, and targeting the products to plastids (Petrasovits et al. 2007) (Table 11.2). PHB accumulated up to 2 % of the leaf dry weight without affecting plant growth and sugar accumulation under greenhouse conditions. Trace amounts of PHB were detected when these enzymes were targeted to the cytosol and PHB was not detected in mitochondrial targeted enzymes. Another industrial compound produced in sugarcane is the aromatic *p*-hydroxybenzoic acid (*p*HBA) (McQualter et al. 2005) (Table 11.2), which is used in the manufacture of polymeric resins and is a natural intermediate in plant and bacterial biosynthetic pathways (Birch 2007). *p*HBA was produced by constitutively expressing the 4-hydroxycinnamoyl-CoA hydratase/lyase (*hchl*) gene from *Pseudomonas fluorescens* under the control of maize *Ubi-1* promoter. The substrate for this enzyme was 4-hydroxycinnamoyl-CoA, a cytosolic phenylpropanoid intermediate. *p*HBA levels accumulated as a glucose conjugate up to 7.3 and 1.5 % of dry weight in leaf and stem tissues, respectively, without detectable phenotypic abnormalities.

Sugarcane has also been used for the production of high-value pharmaceutical proteins. The human cytokine granulocyte macrophage colony stimulating factor (GM-CSF), used for treatment of neutropenia and aplastic anemia, has been successfully produced in sugarcane with identical biological activities to the commercially available protein (Wang et al. 2005a). Expression levels up to 0.02 % of total soluble protein were achieved in leaves by using the constitutive promoters, maize *Ubi-1* or sugarcane *ScUbi9* with the addition of a C-terminal HDEL tag for endoplasmic

reticulum retention of the protein. Field trials that lasted 14 months showed stable expression of the GM-CSF protein with no detrimental effect on plant performance.

4 High Priorities for Future Transformation

Selection of novel traits for engineering as well as optimization of existing technologies and the development of novel systems for efficient transformation of a diverse range of varieties, precise integration and enhanced expression of transgenes, and successful manipulation of metabolic processes would lead to more efficient molecular breeding and the commercial exploitation of biotechnology to improve sugarcane productivity.

4.1 Technologies and Tools

4.1.1 Improved and Novel Transformation Techniques

Biolistic-Mediated Transformation

Genetically engineered sugarcane varieties are mostly developed using particle bombardment for gene delivery (Tables 11.1 and 11.2). Despite its utility, this transformation system has raised environmental safety concerns from the introduction of unwanted vector sequences and antibiotic resistance genes into the plant genome. To address this problem, vector-free minimal transgene cassettes (MGC) have been used to eliminate the use of unnecessary DNA and to produce transgenic plants with low copy number (Lowe et al. 2009; Zhao et al. 2007). In a recent study in sugarcane, an increase of up to 160-fold in the level of transient expression of the EYFP reporter gene has been achieved by using a double terminator MGC (Beyene et al. 2011). The utility of such system needs to be further tested in sugarcane stable transformants.

Agrobacterium-Mediated Transformation

Agrobacterium-mediated gene transfer is gaining more popularity in sugarcane, but is still limited to a few varieties. It is essential to optimize this technology for its application to a broader range of varieties, as it offers excellent advantages, including the possibility of distinct integration of large-sized DNA into the sugarcane genome with low transgene copy insertion and minimal DNA rearrangements.

The efficiency of recovery of transgenic plants using *Agrobacterium*-mediated system in sugarcane has been reported to be relatively low, in the range of 1–5 % (Arencibia and Carmona 2006; Arvinth et al. 2010; Wang et al. 2005b). Approaches

that can significantly improve transformation efficiencies are still under investigation. Browning and necrosis of infected tissues during cocultivation with *Agrobacterium* reduce the competency of cells to be transformed and their recovery after infection (de la Riva et al. 1998; Enríquez-Obregón et al. 1997). The incompatible interaction between *Agrobacterium* and target tissues in monocotyledonous species induces a programmed cell death (PCD) rather than a simple necrosis (Parrot et al 2002; Veena et al. 2003). Transient expression of the animal antiapoptotic genes *Bcl-xL*, *Bcl-2* 3′-untranslated region, and *CED-9* were shown to inhibit up to 90 % of the PCD induced by *Agrobacterium* during cocultivation in banana and sugarcane cell cultures (Khanna et al. 2007), thereby increasing transformation efficiency up to 100- and 35-fold, respectively. Recently identified plant homologs for these anti-apoptotic genes (Doukhanina et al. 2006) could be cointroduced with the trait gene to improve the survival and recovery of sugarcane transformants.

Plastid Transformation

Development of plastid genetic engineering of sugarcane is particularly attractive since it offers a number of unique advantages compared to nuclear transformation technologies, including the precise integration of genes into the plastome by homologous recombination without position effect or gene silencing (Daniell et al. 2005), the expression of polycistronic or multiple genes in a single transcriptional unit (Bock 2001), and the prospect of increased levels of recombinant protein accumulation (De Cosa et al. 2001). Furthermore, transformed plastids are maternally inherited and cannot be transmitted by pollen (Lee et al. 2003), and selectable marker-free transgenic plants can be generated from transformed plastids using homologous recombination (Daniell et al. 2001).

So far, only tobacco chloroplasts can be transformed on a routine basis (Barone et al. 2009; Huang et al. 2002; Lee et al. 2003; Maliga 2004), but substantial progress has been made in expanding the range of this technique to other plant species (reviewed by Warzecha and Henning 2010), including monocotyledonous plants like rice (Lee et al. 2006). So far, attempts to develop plastid transformation in sugarcane have been unsuccessful (Ralph Bock, Max Planck Institute, Germany, personal communication), but future efforts are promising with the availability of the complete nucleotide sequence of the sugarcane chloroplast genome (Asano et al. 2004), and the adoption of vector design and plant transformation and regeneration methods from engineered chloroplasts of other related monocotyledonous plants (Daniell et al. 2005; Warzecha and Henning 2010).

4.1.2 Reduction in the Risk of Transgene Silencing

Transgene silencing continues to be a major impediment in genetically engineered sugarcane (Mudge et al. 2009; Potier et al. 2008a; van der Merwe et al. 2003; Wei et al. 2003). Overcoming this obstacle is a critical requirement for obtaining

functional promoters for practical metabolic engineering and commercialization of transgenic sugarcane. Transgene silencing was shown recently to be prevalent in primary sugarcane transformants, independent of transgene copy number or sites of integration, developmentally regulated, 5′-sequence specific, and initially posttranscriptional (Birch et al. 2010b; Mudge et al. 2009). One of the main strategies to reduce the occurrence of transgene silencing in sugarcane is by cotransformation using genes coding for plant viral suppressor proteins such as PO from *Sugarcane yellow leaf virus* (Mangwende et al. 2009) and P1/HcPro from *Sugarcane mosaic virus* (Anandalakshmi et al. 1998; Zhang et al. 2008) that are now under evaluation in some laboratories (Potier et al. 2008b; Wang et al. 2006, 2007).

The use of constructs that contain matrix (or scaffold) attachment regions (MAR) flanking the transgene expression cassette (Cheng et al. 2001; De Bolle et al. 2007; Petersen et al. 2002; Waldron et al. 2001) may also minimize silencing by protecting the transgene from position effects (Spiker and Thompson 1996). The MAR elements can potentially bind to components of the chromosome scaffold, thereby isolating the intervening DNA from the rest of the chromosome and preventing interference from genes in adjacent regions.

Silencing may also be related to sequence duplication within the transgene cassette, and careful design of transgene constructs may minimize variation in transgene expression and avoid transgene silencing (reviewed by Butaye et al. 2005; De Wilde et al. 2000).

4.1.3 Development of Regulated Promoters

The development of a wider range of novel promoters that differ in their ability to regulate the temporal and spatial expression patterns of transgenes constitute a major priority for the genetic improvement of sugarcane and the production of new products at levels useful for commercialization. The number of promoters available for sugarcane transformation is very limited, and most of them are of constitutive nature (Albert and Wei 2003, 2004; Braithwaite et al. 2004; Christensen et al. 1992; Wei et al. 1999, 2003; Yang et al. 2003). However, when some of these promoters were reintroduced into sugarcane, they were transcriptionally silenced (Wei et al. 2003).

In terms of regulated promoters, a recent report confirmed the isolation of two stem-expressed and stress-inducible promoters for sugarcane *dirigent* and *o-methyltransferase*, putative defense and fiber biosynthesis-related genes (Damaj et al. 2010). These promoters were shown to be functionally active in the stem vascular tissues and are inducible by salicylic acid and jasmonates in leaves and roots of stably transformed sugarcane and rice plants. Three stem- and one meristem-specific promoters have been previously isolated from sugarcane (Abraha 2005; Hansom et al. 1999; Potier and Birch 2001), but their functionality remains to be established.

4.1.4 Availability of New Transformation Vectors

Manipulation of complex metabolic pathways and agronomic traits using genetic engineering most often requires the introduction of multiple genes. This task might not be a limiting factor for transformation using particle bombardment protocols, where up to 13 genes have been successfully introduced in rice (Chen et al. 1998). However, with the increasing use of *Agrobacterium* as a tool of choice for transformation of sugarcane, it is important to use versatile binary systems capable of harboring and transferring multiple genes into the host genome. Two main approaches have been used to deliver multiple T-DNAs either as two sets of T-DNAs in a single vector backbone (Xing et al. 2000) or in two separate binary vectors (Hellens et al. 2000) that can be replicated in the same *Agrobacterium* host. A novel dual binary vector system, pCLEAN, has been developed recently for *Agrobacterium*-mediated plant transformation (Thole et al. 2007). The pCLEAN vectors are based on the pGreen/pSoup system (Hellens et al. 2000), and some of them contain extra *virG* genes in their backbone for enhanced transformation efficiency, and an extra consensus left T-DNA border that reduces the transfer of the backbone sequence into the plant nuclear genome.

Current efforts in sequencing the sugarcane genome will generate a large set of novel genes that will need to be functionally characterized, generating a need for the construction of a suite of vectors for transformation in an efficient and high-throughput manner. Himmelbach et al. (2007) have developed a set of modular binary vectors for cereals to overexpress or knockdown gene expression, under the control of the maize *ubiquitin-1*, the rice *actin 1*, the double *CaMV 35S*, or the wheat epidermis-specific *TaGstA1* promoters. One of the main advantages of this set of vectors is the use of the GATEWAY recombination system (Invitrogen, Carlsbad, California) for its convenience in cloning without the need for compatible restriction sites.

Although *Agrobacterium*-mediated transformation is becoming routine for the production of transgenic sugarcane, there remains some problems associated with this technology. During transformation, the backbone sequences flanking the T-DNA are often integrated into the plant genome together with the bacterial resistance genes, thus creating environmental concerns (De Buck et al. 2000). A recent study has shown that launching T-DNA from the plant-inducible *picA* locus in the *A. tumefaciens* chromosome was found to reduce transgene copy number and almost eliminate the presence of T-DNA backbone sequences in Arabidopsis and maize (Oltmanns et al. 2010). Such a strategy may enable transgenic sugarcane plants produced by *Agrobacterium*-mediated transformation to exhibit higher levels of transgene expression but also free of residual antibiotic resistance genes and backbone sequences.

4.1.5 Targeted Transgene Integration

A new technology, based on the use of gene-targeted zinc-finger nucleases (ZFNs) was developed and used for the generation of organisms and plants with gene-targeted modifications (Porteus 2009). This technology offers the prospect of opening up the

way to enhance targeted integration in sugarcane. ZNFs are engineered DNA-binding proteins that facilitate targeted editing of the genome by creating double-strand breaks in DNA at user-specified locations (Porteus 2009). Double-strand breaks are important for site-specific mutagenesis in that they stimulate the cell's natural DNA-repair processes, namely, homologous recombination and nonhomologous end-joining, resulting in cells with targeted gene deletions, integrations, or modifications. In this regard, substantial increases in the frequency of targeted integration in a broad range of animal and plant model systems have been observed by using this technology (Porteus and Carroll 2005; Townsend et al. 2009; Shukla et al. 2009).

4.1.6 Transgene Stacking

Genetic engineering for complex or combined traits requires the simultaneous expression of multiple transgenes, and is not efficient with the existing transformation techniques. Minichromosome technology provides one solution to the stable expression and maintenance of multiple transgenes in one genome (reviewed by Yu et al. 2007). All the transgenes would reside on an independent minichromosome, not linked to any endogenous genes, and are not subject to position effects. Engineered minichromosomes can now be easily constructed by a telomere-mediated chromosomal truncation strategy after its initial success in maize (Yu et al. 2006). This approach does not rely on the cloning of centromere sequences, and is therefore non-species-specific. An immediate application of the minichromosome technology in sugarcane is to enable stacking of genes involved with herbicide resistance, pest and disease tolerance, and crop quality traits. Entire biochemical pathways can also be added to plants to confer new properties or to synthesize novel metabolites in mass quantities.

4.2 Traits and Products

With the current efforts in sequencing the genome of this crop, genes encoding a whole metabolic pathway could be identified and exploited for engineering of multiple stress traits, or for the production of useful compounds and products.

4.2.1 Improvement of Sucrose Content

Sucrose content is one of the high priority polygenic traits that needs be improved in sugarcane, and several attempts to alter the expression levels of key genes involved in sucrose metabolism were of limited success (reviewed by Lakshmanan et al. 2005), due to the existence of alternative pathways. Progress was mainly achieved by two research groups in increasing either the sucrose content alone (Groenewald and Botha 2001; Wu and Birch 2007) or in enhancing both the sucrose levels and

the cell wall fiber biomass through carbon partitioning in the stem internodes (Groenewald and Botha 2008; van der Merwe et al. 2010). This provides a promising avenue for the future genetic modification of sugar metabolism. Genome sequencing data as well as kinetic studies and structural modeling of sucrose metabolic enzymes will lead to a better understanding of sugar metabolic and transport pathways in order to increase the flux of sugars, and consequently the concentration of sucrose in the storage tissue.

4.2.2 Production of High-Value Sugars

As sugarcane has an abundant supply of stored sucrose in its stem, it can be engineered to produce novel high-value sugars, in particular sucrose metabolites, with promising health benefits. Isomaltulose and its derivatives (Zhang et al. 2002, 2003; Wu and Birch 2005) provide a good example for the production of other potentially functional sugars for human consumption.

4.2.3 Enhancement of Abiotic Stress Tolerance Through "Regulons"

Transcription factors (regulons) that regulate expression of a number of downstream abiotic stress responsive genes (functional genes) represent a more promising approach in the development of abiotic stress tolerant transgenic plants than engineering individual genes (Century et al. 2008; Nakashima and Yamaguchi-Shinozaki 2006). This approach, also called "regulon biotechnology", has been successfully tested in a range of crop plants to confer abiotic stress tolerance. A well-studied family of transcription factors is the Arabidopsis AP2/ERF (APETALA2/ethylene response factor) family, also known as the CBFs/DREB1s (C-repeat binding factors/dehydration-responsive element binding proteins), which is involved in drought, salt, and cold stress tolerance. Overexpression of most of the genes belonging to this group of transcription factors have resulted in single or multiple stress tolerance in a number of plant species (Nakashima et al. 2009; Umezawa et al. 2006; Wang et al. 2003; Zhang et al. 2004).

In sugarcane, overexpression of the drought-induced transcription factor CBF4 gene from Arabidopsis was accompanied by a significant increase in expression of a few assayed genes that are known targets for CBF activation under normal growth conditions, namely, *ERD4* and *p5cs* (McQualter and Dookun-Saumtally 2007). This shows that the CBF4 regulon is at least functional in sugarcane; however, stress tolerance of CBF4 transgenic plants still needs to be established. An AP2/ERF family transcription factor, SoERF3, which is induced in response to ethylene, ABA, salt and wounding has been cloned and characterized from sugarcane (Trujillo et al. 2008). Greenhouse-grown transgenic tobacco plants expressing *SoERF3* driven by a constitutive promoter displayed increased tolerance to drought and osmotic stress (Trujillo et al. 2008). The functionality of this gene in sugarcane is still to be determined.

Reports on regulon biotechnology show the potential and possibilities of adapting and testing predefined sets of regulatory genes from model plants and related grass species. At the same time, a wealth of information that is already available from large-scale transcriptional profiling studies in sugarcane (reviewed by Menossi et al. 2008) could be integrated along with information from sugarcane EST databases, in the process of recruiting regulatory genes for stress tolerance.

4.2.4 Enhancement of Biotic Stress Tolerance

Biotic stress tolerance has already been achieved by introducing single genes into sugarcane (Table 11.1). However, several commercial sugarcane varieties are vulnerable to more than one pathogen or pest. Gene stacking can be used to introduce multiple biotic stress tolerance traits into a single sugarcane variety, and can be exploited in integrated pest management strategies and breeding programs. This is practically feasible with the aid of the advanced modern genetic engineering technologies, and due to the availability of plant- and pathogen-derived biotic stress tolerance genes and antimicrobial peptides (reviewed by Lakshmanan et al. 2005).

4.2.5 Modification of Plant Development

Improvement of Biomass

Improving biomass in sugarcane can be achieved through the genetic manipulation of genes encoding key photosynthetic enzymes, such as the chloroplastic fructose-1,6-bisphosphate that plays a major role in carbon dioxide assimilation and in coordinating carbon and nitrogen metabolism to increase sucrose production (Sahrawy et al. 2004). Manipulation of genes that are involved in the biosynthesis of hormones, such as the brassinosteroids (Sakamato et al. 2006), as well as engineering for delayed flowering may have a great impact on enhancing the sugarcane biomass.

Modulation of Flowering

Inhibition of flowering has considerable economic significance in commercial sugarcane production. Flower production triggers a senescence process during which the stored sucrose in the sugarcane stem starts to be converted into reducing sugars, diminishing significantly the recovery of sucrose during the extraction process. The lack of success in transferring natural traits of no tassel formation from *Saccharum* species like *S. edule* by conventional breeding (Premachandran 2006), as well as the undesirable side effects caused by treatment with growth regulators like Ethephon (Chong et al. 2010; Donaldson 1996; Moore and Osgood 1989) make the use of genetic engineering more appealing for modulating flowering in sugarcane. Most of the major genes controlling flowering have been identified and

functionally characterized in dicotyledonous species like Arabidopsis (reviewed by Kaufmann et al. 2010; Putterill 2001); however, very little is known about the genes controlling this complex trait in monocotyledonous plants. A homolog for the *Arabidopsis* LEAFY (*AtLFY*) gene in sugarcane has been recently identified from the SUCEST database by Ulian (2006). Expression of the LFY gene from *S. officinarum* (*SoLFY*) in antisense orientation was sufficient to suppress flowering in some of the transgenic sugarcane lines tested under field conditions (Ulian 2006). Other studies in tobacco have shown that expression of the Arabidopsis flower repressor, *FLOWERING LOCUS C*, enhanced biomass production (Salehi et al. 2005). Transgenic sugarcane lines with impaired flowering and high biomass would be a valuable source of cellulosic biofuel. They can also be useful as target tissue in engineering sugarcane for novel agronomic traits to prevent the vertical and horizontal flow of transgenes.

4.2.6 Production of High-Value Alternative Products and Proteins

Sugarcane, as a plant system, needs to be optimized through genetic engineering to meet the demand for renewable biomaterials and high-value proteins. Optimal amounts of sucrose and value-added forms of multiple natural constituents can be coproduced, considering the capacity of sugarcane to accumulate a large biomass and high sucrose levels.

Sugarcane can be potentially engineered using genes that target the following products and proteins:

1. Modified lignocellulosic fiber. The production of cellulosic ethanol from sugarcane residues and high biomass-producing energycane varieties can be enhanced by genetically modifying the cell wall composition or de-polymerizing the complex carbohydrates. Approaches may aim at reducing lignin content by down-regulation of genes involved in lignin biosynthesis (Chen and Dixon 2007; Hisano et al. 2009; Ralph et al. 2006; Simmons et al. 2010), or by expressing thermostable cellulases and other hydrolases of biomass substrates (Sainz 2009; Sticklen 2006, 2008; Taylor et al. 2008) to reduce the need for commercial enzymatic hydrolysis and pretreatment. Plant biomass can also be converted into fermentable glucose for ethanol production following expression of endoglucanases as reported in rice (Oraby et al. 2007; Sticklen 2006).
2. Proteins with therapeutical applications. As plant expression platforms continue to improve in terms of posttranslational modifications and overall protein yield, more sugarcane-made proteins for medical use are likely to become economically feasible in the future. Proteins expressed in sugarcane can include vaccines, insulin, interferons and pharmaceuticals that have already been produced in different plant species (reviewed by Shiermeyer and Shillberg 2010).
3. Sucrose-derived polymers such as starch, fructans, tagatose, and xylitol with important health impact (reviewed by Brumbley et al. 2008).

4. Aromatic compounds derived from the aromatic and wax biosynthetic pathways in sugarcane (reviewed by Birch 2007).
5. Bioplastic precursors (reviewed by Birch 2007).
6. Industrial enzymes (Howard and Hood 2005).
7. Proteins with valuable fibrous or adhesive properties (Scheller and Conrad 2005).

5 Constraints

The lack of efficient promoter expression systems, the phenomenon of transgene silencing, and the limited information on the inheritance and segregation pattern of transgenes in sugarcane are considered as immediate constraints. There are also considerable safety measures as well as regulatory and public perception issues related to field-grown genetically modified (GM) sugarcane that need to be addressed.

5.1 Factors Affecting Transgene Expression and Performance

5.1.1 Suitable Promoters

A wide range of strongly regulated promoters in terms of tissue-specificity and inducibility is not yet available to meet the critical needs of metabolic engineering and renewable biomaterial and biofuel production in sugarcane. A detailed functional analysis of the potentially existing promoters in driving transgene expression at different stages of sugarcane development is also lacking.

5.1.2 Transgene Inheritance and Silencing

The problem of efficient transgene silencing in sugarcane complicates the development of functional constitutive and regulated promoter systems associated with useful endogenous expression patterns (Mudge et al. 2009), as well as the portability of efficient heterologous promoters into sugarcane (Potier et al. 2008a). Transgene silencing is currently one of the major limiting factors to produce a large number of transgenic sugarcane varieties, and to achieve commercially useful production levels and transgene stability. Transgene position effects and silencing mechanisms can alter levels of expression and its stability across time. A major concern for commercialization of transgenic sugarcane is the instability in transgene expression and consequent phenotypic variability in consecutive generations in the field, especially for transgenic plants carrying multiple copies of transgenes. It is therefore crucial that rigorous field agronomic analyses and trait evaluations be conducted across several generations to ensure transgene stability.

5.1.3 Somaclonal Variation

Several studies in sugarcane have revealed the existence of cell and tissue culture-induced somaclonal variation (Rajeswari et al. 2009) as well as transformation-associated variation during biolistic (Gilbert et al. 2005, 2009; Vickers et al. 2005b) and *Agrobacterium*-mediated transformations (Carmona et al. 2005). Significant differences in agronomic performance of the transformed plants compared to the nontransformed parental variety were reported (Vickers et al. 2005b) and different genotyping techniques that included amplified fragment length polymorphism (Carmona et al. 2005), simple sequence repeat (Gilbert et al. 2009) and random amplification of polymorphic DNA (Hussain 2005) were used to detect changes at the genomic level. This indicates the need for further development of less aggressive transformation techniques and shorter exposure of transformed explants to in vitro culture conditions.

5.2 Gene Flow

In general, sugarcane offers a high level of transgene containment. However, understanding the sexual reproductive behavior of sugarcane under different growing conditions is required to determine possible environmental consequences of GM plants. Such studies are specially lacking under commercial settings as sugarcane is propagated vegetatively, and many of the varieties used do not produce viable pollen. Recently, a generic framework has been developed to address the risk arising from nonintentional gene escape from GM sugarcane (Olivares-Villegas et al. 2010). This framework identifies the potential environmental effects resulting from the sexual propagation of sugarcane in commercial fields, including the potential to flower in the target environment, the presence of species that are sexually compatible with sugarcane, and the production of fertile seed and its germination and establishment. Such an initiative is among the first ones to address one of the neglected but indispensable criteria required by regulatory authorities for the release of GM sugarcane.

5.3 Commercialization

A wide range of novel transgenic sugarcane varieties have been generated and tested in field trials (Tables 11.1 and 11.2), but no commercial transgenic variety has yet been released. However, news releases from plant biotechnology companies indicate that commercialization is on the horizon. Scientific, legislative, and public perception issues will then need to be addressed extensively for the successful release of transgenic sugarcane.

5.4 Regulatory and Public Perception Issues

The shortage of inexpensive crude oil due to rapid industrial development, political instability in major crude oil producing regions and reductions in oil reserves has resulted in a transition from a petroleum-based economy to a bio-based one. This change has been brought about by the private sector investing heavily into research on biofuels, primarily to produce ethanol. The production and release of transgenic sugarcane has the potential to elevate the yield of biofuels by modifying the crop to increase sugar and biomass per hectare, utilize less water and fertilizer, resist pests and diseases, grow on marginal land and tolerate relatively dry and cold climates. However, are there constraints on the release of genetically modified sugarcane to both human health and the environment? For more than 20 years, over 400 million hectares has been used for the commercial release of transgenic crops and if transgenes are safe, then the GM-product is also safe (Bradford et al. 2005). However, environmental campaigners along with public perception have offered resistance to field trials of GM crops despite there being no scientific foundation for concern (Miller 2007). To date, some large companies, which buy raw sugar, are only accepting GM-free material. In terms of the future, it is likely that transgenic sugarcane for refined sugar production will be accepted into the market, as the product is free from foreign genes or gene products (Birch and Maretzki 1993; Taylor et al. 1999). Nevertheless, for transgenic sugarcane to be utilized in the pharmaceutical industry, clear guidelines need to be integrated to cover any problematic issues that could emerge. For example, sugarcane producing bioplastics such as *p*-hydrobenzoic acid (*p*HBA) and polyhydroxybutyrate (PHBs), which in nature are used as carbon sinks, are completely biodegradable and offer no danger to human health or the environment. At present, there have been limited reports on the release of genetically modified sugarcane to the field but it is clear through the rapid development of biotechnology that this will escalate by the heavy investment from the private sector.

References

Abraha TG (2005) Isolation and characterization of a culm-specific promoter element from sugarcane. MSc dissertation, Department of Botany and Zoology, Institute of Biotechnology, Stellenbosch University, South Africa

Albert HH, Wei H (2003) Promoter of the sugarcane UBI4 gene. US Patent No. 6,638,766

Albert HH, Wei H (2004) Sugarcane *ubi9* gene promoter sequence and methods of use thereof. US Patent No. 6,686,513

Altpeter F, Oraby H (2010) Sugarcane. In: Kempken F, Jung C (eds) Genetic modification of plants Biotechnology in agriculture and forestry, vol 64. Springer-Verlag, Berlin, pp 453–472

Altpeter F, Baisakh N, Beachy R, Bock R, Capell T, Christou P, Daniell H, Datta K, Datta S, Dix PJ, Fauquet C, Huang N, Kohli A, Mooibroek H, Nicholson L, Nguyen TT, Nugent G, Raemakers K, Romano A, Somers DA, Stoger E, Taylor N, Visser R (2005) Particle bombardment and the genetic enhancement of crops: myths and realities. Mol Breed 15:305–327

Anandalakshmi R, Pruss GJ, Ge X, Marathe R, Mallory AC, Smith TH, Vance VB (1998) A viral suppressor of gene silencing in plants. Proc Natl Acad Sci USA 95:13079–13084

Arencibia AD, Carmona ER (2006) Sugarcane (*Saccharum* spp.). In: Wang K (ed) Methods in molecular biology, vol 344, 2nd edn. *Agrobacterium* protocols, vol 2. Humana Press Inc., New Jersey, pp 227–235

Arencibia A, Molina P, Gutiérrez C, Fuentes A, Greenidge V, Menéndez E, De la Riva G, Selman-Houssein G (1992) Regeneration of transgenic sugarcane (*Saccharum officinarum* L.) plants from intact meristematic tissues transformed by electroporation. Biotechnol Aplicada 9:156–165

Arencibia A, Molina PR, de la Riva G, Selman-Housein G (1995) Production of transgenic sugarcane (*Saccharum officinarum* L) plants by intact cell electroporation. Plant Cell Rep 14:305–309

Arencibia A, Vázquez RI, Prieto D, Téllez P, Carmona ER, Coego A, Hernández L, de la Riva GA, Selman-Housein G (1997) Transgenic sugarcane plants resistant to stem borer attack. Mol Breed 3:247–255

Arencibia AD, Carmona ER, Téllez P, Chan M-T, Yu S-M, Trujillo LE, Oramas P (1998) An efficient protocol for sugarcane (*Saccharum* spp. L.) transformation mediated by *Agrobacterium tumefaciens*. Transgenic Res 7:213–222

Arencibia AD, Carmona ER, Cornide MT, Castiglione S, O'Relly J, Chinea A, Oramas P, Sala F (1999) Somaclonal variation in insect-resistant sugarcane (*Saccharum* hybrid) plants produced by cell electroporation. Transgenic Res 8:349–360

Arvinth S, Arun S, Selvakesavan RK, Srikanth J, Mukunthan N, Kumar PA, Premachandran MN, Subramonian N (2010) Genetic transformation and pyramiding of aprotinin-expressing sugarcane with *cry1Ab* for shoot borer (*Chilo infuscatellus*) resistance. Plant Cell Rep 29:383–395

Asano T, Tsudzuki T, Takahashi S, Shimada H, Kadowaki KI (2004) Complete nucleotide sequence of the sugarcane (*Saccharum officinarum*) chloroplast genome: a comparative analysis of four monocot chloroplast genomes. DNA Res 11:93–99

Barba R, Nickell LG (1969) Nutrition and organ differentiation in tissue cultures of sugarcane, a monocotyledon. Planta 89:299–302

Barone P, Zhang X-H, Widholm JM (2009) Tobacco plastid transformation using the feed-back-insensitive anthranilate synthase [α]-subunit of tobacco (ASA2) as a new selectable marker. J Exp Bot 60:3195–3202

Beck E, Ludwig G, Auerswald EA, Reiss B, Schaller H (1982) Nucleotide sequence and exact localization of the neomycin phosphotransferase gene from transposon Tn5. Gene 9:327–336

Behera KK, Sahoo S (2009) Rapid *in vitro* micro propagation of sugarcane (*Saccharum officinarum* L. cv-Nayana) through callus culture. Nat Sci 7:1–10

Beyene G, Buenrostro-Nava MT, Damaj MB, Gao S, Molina J, Mirkov TE (2011) Unprecedented enhancement of reporter gene expression from minimal cassettes using a double terminator. Plant Cell Rep 30:13–25

Birch RG (2007) Metabolic engineering in sugarcane: assisting the transition to a bio-based economy. In: Verpoorte RA, Alfermann AW, Johnson TS (eds) Applications of plant metabolic engineering. Springer, Berlin, pp 249–281

Birch RG, Franks T (1991) Development and optimization of microprojectile systems for plant genetic transformation. Aust J Plant Physiol 18:453–469

Birch RG, Maretzki A (1993) Transformation of sugarcane. In: Bajaj YPS (ed) Plant proto-plasts and genetic engineering IV. Biotechnology in agriculture and forestry, vol 23, Springer, Heidelberg, pp 248–360

Birch GR, Shen B, Sawyer BJB, Huttner E, Tucker WQJ, Betzner AS (2010a) Evaluation and application of a luciferase fusion system for rapid *in vivo* analysis of RNAi targets and constructs in plants. Plant Biotechnol J 8:465–475

Birch RG, Bower RS, Elliott AR (2010b) Highly efficient, 5′-sequence-specific transgene silencing in a complex polyploid. Trop Plant Biol 3:75–87

Bock R (2001) Transgenic plastids in basic research and plant biotechnology. J Mol Biol 312:425–438

Botha FC, Sawyer BJB, Birch RG (2001) Sucrose metabolism in the culm of transgenic sugarcane with reduced soluble acid invertase activity. Proc Int Soc Sugar Cane Technol 24:588–591

Bower R, Birch RG (1992) Transgenic sugarcane plants via microprojectile bombardment. Plant J 2:409–416

Bower R, Elliott AR, Potier BAM, Birch RG (1996) High-efficiency, microprojectile-mediated cotransformation of sugarcane, using visible or selectable markers. Mol Breed 2:239–249

Bradford KJ, Deynze AV, Gutterson N, Parrott W, Strauss SH (2005) Regulating transgenic crops sensibly: lessons from plant breeding, biotechnology and genomics. Nat Biotechnol 23:439–444

Braga DPV, Arrigoni EDB, Silva-Filho MC, Ulian EC (2003) Expression of the Cry1Ab protein in genetically modified sugarcane for the control of *Diatraea saccharalis* (Lepidoptera: Crambidae). J New Seeds 5:209–222

Braithwaite KS, Geijskes RJ, Smith GR (2004) A variable region of the sugarcane bacilliform virus (SCBV) genome can be used to generate promoters for transgene expression in sugarcane. Plant Cell Rep 23:319–326

Brasileiro ACM, Aragão FJL (2001) Marker genes for *in vitro* selection of transgenic plants. J Plant Biotechnol 3:113–121

Brumbley SM, Snyman SJ, Gnanasambandam A, Joyce P, Hermann SR, da Silva JAG, McQualter RB, Wang ML, Egan BT, Patterson AH, Albert HH, Moore PH (2008) Sugarcane. In: Kole C, Hall TC (eds) Compendium of transgenic crop plants: transgenic sugar, tuber and fiber crops. Wiley-Blackwell, Oxford, pp 1–58

Butaye KMJ, Cammue BPA, Delauré SL, De Bolle MFC (2005) Approaches to minimize variation of transgene expression in plants. Mol Breed 16:79–91

Carmona ER, Arencibia AD, Lopez J, Simpson J, Vargas D, Sala F (2005) Analysis of genomic variability in transgenic sugarcane plants produced by *Agrobacterium tumefaciens* infection. Plant Breed 124:33–38

Century K, Reuber TL, Ratcliffe OJ (2008) Regulating the regulators: the future prospects for transcription-factor-based agricultural biotechnology products. Plant Physiol 147:20–29

Chen F, Dixon RA (2007) Lignin modification improves fermentable sugar yields for biofuel production. Nat Biotechnol 25:759–761

Chen WC, Gartland KMA, Davey MR, Sotak R, Gartland JS, Mulligan BJ, Power JB, Cocking EC (1987) Transformation of sugarcane protoplasts by direct uptake of a selectable chimeric gene. Plant Cell Rep 6:297–301

Chen WH, Davey MR, Power JB, Cocking EC (1988) Control and maintenance of plant regeneration in sugarcane callus cultures. J Exp Bot 39:251–261

Chen L, Marmey P, Taylor NJ, Brizard J, Espinoza C, D'Cruz P, Huet H, Zhang S, de Kochko A, Beachy RN, Fauquet CM (1998) Expression and inheritance of multiple transgenes in rice plants. Nat Biotechnol 16:1060–1064

Cheng Z, Targolli J, Wu R (2001) Tobacco matrix attachment region sequence increased transgene expression levels in rice plants. Mol Breed 7:317–327

Chong BF, Bonnett GD, Glassop D, O'Shea MG, Brumbley SM (2007) Growth and metabolism in sugarcane are altered by the creation of a new hexose-phosphate sink. Plant Biotechnol J 5:240–253

Chong BF, Mills E, Bonnett GD, Gnanasambandam A (2010) Early exposure to ethylene modifies shoot development and increases sucrose accumulation rate in sugarcane. J Plant Growth Regul 29:149–163

Chowdhury MKU, Vasil IK (1992) Stably transformed herbicide resistance callus of sugarcane via microprojectile bombardment of cell suspension cultures and electroporation of protoplasts. Plant Cell Rep 11:494–498

Christensen AH, Sharrock RA, Quail PH (1992) Maize polyubiquitin genes: structure, thermal perturbation of expression and transcript splicing, and promoter activity following transfer to protoplasts by electroporation. Plant Mol Biol 18:675–689

Christy LA, Arvinth S, Saravanakumar M, Kanchana M, Mukunthan N, Srikanth J, George T, Subramonian N (2009) Engineering sugarcane cultivars with bovine pancreatic trypsin inhibitor

(aprotinin) gene for protection against top borer (*Scripophaga excerptalis* Walker). Plant Cell Rep 28:175–184

D'Hont A, Ison D, Alix K, Roux C, Glaszmann JC (1998) Determination of basic chromosome numbers in the genus *Saccharum* by physical mapping of ribosomal RNA genes. Genome 41:221–225

D'Hont A, Souza GM, Menossi M, Vincentz M, Van-Sluys M-A, Glaszmann JC, Ulian E (2008) Sugarcane: a major source of sweetness, alcohol, and bioenergy. In: Moore PH, Ming R (eds) Tropical crop plant genomics. Springer-Verlag, New York, pp 483–513

Dai SH, Zheng P, Marmey P, Zhang SP, Tian WZ, Chen SY, Beachy RN, Fauquet C (2001) Comparative analysis of transgenic rice plants obtained by *Agrobacterium*-mediated transformation and particle gun bombardment. Mol Breed 7:25–33

Damaj MB, Kumpatla SP, Emani C, Beremand PD, Reddy AS, Rathore KS, Buenrostro-Nava MT, Curtis IS, Thomas TL, Mirkov TE (2010) Sugarcane *DIRIGENT* and *O-METHYLTRANSFERASE* promoters confer stem-regulated gene expression in diverse monocots. Planta 231:1439–1458

Daniell H, Muthukumar B, Lee SB (2001) Marker free transgenic plants: engineering the cholroplast genome without the use of antibiotic selection. Curr Genet 39:109–116

Daniell H, Kumar S, Dufourmantel N (2005) Breakthrough in chloroplast genetic engineering of agronomically important crops. Trends Biotechnol 23:238–245

Darbani B, Elimanifar A, Stewart CN, Camargo WN (2007) Methods to produce marker-free transgenic plants. Biotechnol J 2:83–90

De Block M, Botterman J, Vandewiele M, Dockx J, Thoen C, Gossele V, Movva NR, Thompson C, Van Montague M, Leemans J (1987) Engineering herbicide resistance in plants by expression of a detoxifying enzyme. EMBO J 6:2513–2518

De Bolle MF, Butaye KM, Goderis IJ, Wouters PF, Jacobs A, Delaure SL, Depicker A, Cammue BP (2007) The influence of matrix attachment regions on transgene expression in *Arabidopsis thaliana* wild type and gene silencing mutants. Plant Mol Biol 63:533–543

De Buck S, de Wilde C, van Montagu M, Depicker A (2000) T-DNA vector backbone sequences are frequently integrated into the genome of transgenic plants obtained by *Agrobacterium*-mediated transformation. Mol Breed 6:459–468

De Cosa B, Moar W, Lee SB, Miller M, Daniell H (2001) Overexpression of the *Bt cry2*Aa2 operon in chloroplasts leads the formation of insecticidal crystals. Nat Biotechnol 19:71–74

de la Riva GA, González-Cabrera J, Vázquez-Padrón R, Ayra-Pardo C (1998) *Agrobacterium tumefaciens*: a natural tool for plant transformation. Electr J Biotechnol 1:118–133

De Wilde C, Van Houdt H, De Buck S, Angenon G, Jarger GD, Depicker A (2000) Plants as bioreactors for protein production: avoiding the problem of transgene silencing. Plant Mol Biol 43:347–359

Desai NS, Suprasanna P, Bapat VA (2004) Simple and reproducible protocol for direct somatic embryogenesis from cultured immature inflorescence segments of sugarcane (*Saccharum* spp.). Curr Sci 87:764–768

Donaldson RA (1996) Effects of ethephon applied to two sugarcane varieties to prevent flowering. Proc S Afr Sug Technol Assoc 70:38–41

Doukhanina EV, Chen S, van der Zalm E, Godzik A, Reed J, Dickman MB (2006) Identification and functional characterization of the BAG protein family in *Arabidopsis thaliana*. J Biol Chem 281:18793–18801

Elliott AR, Campbell JA, Bretell RIS, Grof CPL (1998) *Agrobacterium*-mediated transformation of sugarcane using GFP as a screenable marker. Aust J Plant Physiol 25:739–743

Enríquez-Obregón GA, Vázquez-Padrón RI, Prieto-Samsonov DL, Perez M, Selman-Housein G (1997) Genetic transformation of sugarcane by *Agrobacterium tumefaciens* using anti-oxidant compounds. Biotechnol Aplicada 14:169–174

Enríquez-Obregón GA, Vázquez-Padrón RI, Prieto-Samsonov DL, De la Riva GA, Selman-Housein G (1998) Herbicide-resistant sugarcane (*Saccharum officinarum* L.) plants by *Agrobacterium*-mediated transformation. Planta 206:20–27

US Environmental Protection Agency (1994) Neomycin phosphotransferase: II. Tolerance exemption. Fed Regist 56:4935

Falco MC, Silva-Filho MC (2003) Expression of soybean proteinase inhibitors in transgenic sugarcane plants: effects on natural defense against *Diatrae saccharalis*. Plant Physiol Biochem 41:761–766

Falco MC, Tulmann Neto A, Ulian EC (2000) Transformation and expression of a gene for herbicide resistance in Brazilian sugarcane. Plant Cell Rep 19:1188–1194

Fitch MMM, Lehrer AT, Komor E, Moore PH (2001) Elimination of sugarcane yellow leaf virus from infected sugarcane plants by meristem tip culture visualized by tissue immunoassay. Plant Pathol 50:676–680

Franks T, Birch RG (1991) Gene transfer into intact sugarcane cells using microprojectile bombardment. Aust J Plant Physiol 18:471–480

Gallo-Meagher M, Irvine J (1993) Effects of tissue type and promoter strength on transient GUS expression in sugarcane following particle bombardment. Plant Cell Rep 12:666–670

Gallo-Meagher M, Irvine JE (1996) Herbicide resistant sugarcane containing the *bar* gene. Crop Sci 36:1367–1374

Gambley RL, Ford R, Smith GR (1993) Microprojectile transformation of sugarcane meristems and regeneration of shoots expressing *β-glucuronidase*. Plant Cell Rep 12:343–346

Gambley RL, Bryant JD, Masel NP, Smith GR (1994) Cytokinin-enhanced regeneration of plants from microprojectile bombarded sugarcane meristematic tissue. Aust J Plant Physiol 21:603–612

Gilbert RA, Gallo-Meagher M, Comstock JC, Miller JD, Jain M, Abouzid A (2005) Agronomic evaluation of sugarcane lines transformed for resistance to sugarcane mosaic virus strain E. Crop Sci 45:2060–2067

Gilbert RA, Glynn NC, Comstock JC, Davis MJ (2009) Agronomic performance and genetic characterization of sugarcane transformed for resistance to sugarcane yellow leaf virus. Field Crops Res 111:39–46

Gnanasambandam A, Birch RG (2004) Efficient developmental mistargeting by the sporamin NTPP vacuolar signal to plastids in young leaves of sugarcane and Arabidopsis. Plant Cell Rep 23:435–447

Gnanasambandam A, Polkinghorne IG, Birch RG (2007) Heterologous signals allow efficient targeting of a nuclear-encoded fusion protein to plastids and endoplasmic reticulum in diverse plant species. Plant Biotechnol J 5:290–296

Goldstein DA, Tinland B, Gilbertson LA, Staub JM, Bannon GA, Goodman RE, McCoy RL, Silvanovich A (2005) Human safety and genetically modified plants: a review of antibiotic resistance markers and future transformation selection technologies. J Appl Microbiol 99:7–23

Groenewald J-H, Botha FC (2001) Down regulating pyrophosphate-dependent phosphofructokinase (PFP) in sugarcane. Proc Int Soc Sugar Cane Technol 24:592–594

Groenewald J-H, Botha FC (2008) Down-regulation of pyrophosphate: fructose 6-phosphate 1-phosphotransferase (PFP) activity in sugarcane enhances sucrose accumulation in immature internodes. Transgenic Res 17:85–92

Hansom S, Bower R, Zhang L, Potier B, Elliot A, Basnayake S, Cordeiro B, Hogarth DM, Cox M, Berding N, Birch RG (1999) Regulation of transgene expression in sugarcane. Proc Int Soc Sugar Cane Technol 23:278–289

Hauptmann RM, Vasil V, Ozaias-Aikins P, Tabaeizadeh Z, Rogers SG, Fraley RT, Horsch RB, Vasil IK (1988) Evaluation of selectable markers for obtaining stable transformants in the Gramineae. Plant Physiol 86:602–606

Heinz DJ, Mee GWP (1969) Plant differentiation from callus tissue of *Saccharum* species. Crop Sci 9:346–348

Hellens RP, Edwards EA, Leyland NR, Bean S, Mullineaux PM (2000) pGreen: a versatile and flexible binary Ti vector for *Agrobacterium*-mediated plant transformation. Plant Mol Biol 42:819–832

Himmelbach A, Zierold U, Hensel G, Riechen J, Douchkov D, Schweizer P, Kumlehn J (2007) A set of modular binary vectors for transformation of cereals. Plant Physiol 145:1192–2000

Hisano H, Nandakumar R, Wang Z-Y (2009) Genetic modification of lignin biosynthesis for improved biofuel production. In Vitro Cell Dev Biol Plant 45:306–313

Ho W-J, Vasil IK (1983) Somatic embryogenesis in sugarcane (*Saccharum offcinarum* L.). The morphology and ontogeny of somatic embryos. Protoplasma 118:169–180

Howard JA, Hood EE (2005) Bioindustrial and biopharmaceutical products produced in plants. Adv Agron 85:91–124

Huang FC, Klaus SM, Herz S, Zou Z, Koop HU, Golds TJ (2002) Efficient plastid transformation in tobacco using the *aphA-6* gene and kanamycin selection. Mol Genet Genomics 268:19–27

Hussain A (2005) Biochemical and molecular investigation of somaclonal variants in sugarcane (*Saccharum officinarum* L. cv CoL-54). PhD dissertation, School of Biological Sciences, University of the Punjab, Lahore, Pakistan

Ingelbrecht IL, Irvine JE, Mirkov TE (1999) Post-transcriptional gene silencing in transgenic sugarcane. Dissection of homology-dependent virus resistance in a monocot that has a complex polyploid genome. Plant Physiol 119:1187–1197

Irvine JE, Benda GTA (1985) Sugarcane mosaic virus in plantlets regenerated from diseased leaf tissue. Plant Cell Tissue Org Cult 5:101–106

Irvine JE, de Almedia CG (1991) Delivery of plasmid DNA through particle bombardment with an airless sprayer. Texas A&M Misc Publ 1726

Jain M, Chengalrayan K, Abouzid A, Gallo M (2007) Prospecting the utility of a PMI/mannose selection system for the recovery of transgenic sugarcane (*Saccharum* spp. hybrid) plants. Plant Cell Rep 28:581–590

Joyce PA, McQualter RB, Bernard MJ, Smith GR (1998) Engineering for resistance to SCMV in sugarcane. Acta Hort 461:385–391

Joyce P, Kuwahata M, Turner N, Lakshmanan P (2010) Selection system and co-cultivation medium are important determinants of *Agrobacterium*-mediated transformation of sugarcane. Plant Cell Rep 29:173–183

Kaufmann K, Wellmer F, Muiño JM, Ferrier T, Wuest SE, Kumar V, Serrano-Mislata A, Madueño F, Krajewski P, Meyerowitz EM, Angenent GC, Riechmann JL (2010) Orchestration of floral initiation by APETALA1. Science 328:85–89

Khan SA, Rashid H, Chaudhary MF, Chaudhry Z, Afroz A (2008) Rapid micropropagation of three elite sugarcane (*Saccharum officinarum* L.) varieties by shoot tip culture. Afr J Biotechnol 7:2174–2180

Khanna HK, Paul JY, Harding RM, Dickman MB, Dale JL (2007) Inhibition of *Agrobacterium*-induced cell death by antiapoptotic gene expression leads to very high transformation efficiency of banana. Mol Plant Microbe Interact 20:1048–1054

Lakshmanan P, Geijskes RJ, Aitken KS, Grof CLP, Bonnett GD, Smith GR (2005) Sugarcane biotechnology: the challenges and opportunities. In Vitro Cell Dev Biol Plant 41:345–363

Lakshmanan P, Geijskes RJ, Wang LF, Elliott A, Grof CPI, Smith GR (2006) Developmental and hormonal regulation of direct shoot organogenesis and somatic embryogenesis in sugarcane (*Saccharum* spp. interspecific hybrids) leaf culture. Plant Cell Rep 25:1007–1015

Lee TSG (1987) Micropropagation of sugarcane (*Saccharum* spp.). Plant Cell Tissue Org Cult 10:47–55

Lee S-B, Kwon H-B, Kwon S-J, Park S-C, Jeong M-J, Han S-E, Byun M-O, Daniell H (2003) Accumulation of trehalose within transgenic chloroplasts confers drought tolerance. Mol Breed 11:1–13

Lee SM, Kang K, Chung H, Yoo SH, Xu XM, Lee SB, Cheong JJ, Daniell H, Kim M (2006) Plastid transformation in the monocotyledonous cereal crop, rice (*Oryza sativa*) and transmission of transgenes to their progeny. Mol Cells 21:401–410

Legaspi JC, Mirkov TE (2000) Evaluation of transgenic sugarcane against stalk borers. Proc Int Soc Sugar Cane Technol 4:68–71

Leibbrandt NB, Snyman SJ (2003) Stability of gene expression and agronomic performance of a transgenic herbicide-resistant sugarcane line in South Africa. Crop Sci 43:671–677

Liu DW, Oard SV, Oard JH (2003) High transgene expression levels in sugarcane (*Saccharum officinarum* L.) driven by the rice ubiquitin promoter RUBQ2. Plant Sci 165:743–750

Lowe BA, Shiva Prakash N, Way M, Mann MT, Spencer TM, Boddupalli RS (2009) Enhanced single copy integration events in corn via particle bombardment using low quantities of DNA. Transgenic Res 18:831–840

Ma H, Albert HH, Paull R, Moore PH (2000) Metabolic engineering of invertase activities in different subcellular compartments affects sucrose accumulation in sugarcane cells. Aust J Plant Physiol 27:1021–1030

Maliga P (2004) Plastid transformation in higher plants. Annu Rev Plant Biol 55:289–313

Mangwende T, Wang ML, Borth W, Hu J, Moore PH, Mirkov TE, Albert HH (2009) The P0 gene of *Sugarcane yellow leaf virus* encodes an RNA silencing suppressor with unique activities. Virology 384:38–50

Manickavasagam M, Ganapathi A, Anbazhagan VR, Sudhakar B, Selvaraj N, Vasudevan A, Kasthurirengan S (2004) *Agrobacterium*-mediated genetic transformation and development of herbicide-resistant sugarcane (*Saccharum* species hybrids) using axillary buds. Plant Cell Rep 23:134–143

McQualter RB, Dookun-Saumtally A (2007) Expression profiling of abiotic-stress inducible genes in sugarcane. Proc Int Soc Sugar Cane Technol 26:878–888

McQualter RB, Dale JL, Harding RM, McMahon JA, Smith GR (2004) Production and evaluation of transgenic sugarcane containing a *Fiji disease virus* (FDV) genome segment S9-derived synthetic resistance gene. Aust J Agric Res 55:139–145

McQualter RB, Chong BF, Meyer K, Van Dyk DE, O'Shea MG, Walton NJ, Viitanen PV, Brumbley SM (2005) Initial evaluation of sugarcane as a production platform for *p*-hydroxybenzoic acid. Plant Biotechnol J 3:29–41

Menossi M, Silva-Filho MC, Vincentz M, Van-Sluys MA, Souza GM (2008) Sugarcane functional genomics: gene discovery for agronomic trait development. Int J Plant Genomics 458732. doi:10.1155/2008/458732

Miller H (2007) Biotech's defining moments. Trends Biotechnol 25:56–59

Ming R, Moore PH, Wu KK, D'Hont A, Glaszmann JC, Tew TL, Mirkov TE, da Silva J, Jifon J, Rai M, Schnell RJ, Brumbley SM, Lakshmanan P, Comstock JC, Paterson AH (2006) Sugarcane improvement through breeding and biotechnology. In: Janick J (ed) Plant breeding reviews, vol 27, John Wiley & Sons, Inc., pp 15–118

Molinari HBC, Marur CJ, Daros E, Campos MKF, Carvalho JFRP, Bespalhok Filho JC, Pereira LFP, Vieira LGE (2007) Evaluation of the stress-inducible production of proline in transgenic sugarcane (*Saccharum* spp.): osmotic adjustment, chlorophyll fluorescence and oxidative stress. Physiol Plant 130:218–229

Moore PH (2005) Integration of sucrose accumulation processes across hierarchical scales: towards developing an understanding of the gene-to-crop continuum. Field Crops Res 92:119–135

Moore PH, Osgood RV (1989) Prevention of flowering and increasing sugar yield of sugarcane by application of ethephon (2-chloroethylphosphonic acid). J Plant Growth Regul 8:205–210

Mudge SR, Osabe K, Casu RE, Bonnett GD, Manners JM, Birch RG (2009) Efficient silencing of reporter transgenes coupled to known functional promoters in sugarcane, a highly polyploid crop species. Planta 229:549–558

Murashige T, Skoog F (1962) A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol Plant 15:473–497

Nakashima K, Yamaguchi-Shinozaki Y (2006) Regulons involved in osmotic stress-responsive and stress-responsive gene expression in plants. Physiol Plant 126:62–71

Nakashima K, Ito Y, Yamaguchi-Shinozaki K (2009) Transcriptional regulatory networks in response to abiotic stresses in Arabidopsis and grasses. Plant Physiol 149:88–95

Nutt KA, Allsopp PC, McGhie TK, Shepherd KM, Joyce PA, Tayoor GO, McQualter RB, Smith GR (1999) Transgenic sugarcane with increased resistance to canegrubs. Proc Aust Soc Sugar Cane Technol 21:171–17630

Olivares-Villegas JJ, Berding N, Morgan T, Bennett GD (2010) A support framework for deployment of genetically modified sugarcane: identifying potential risks from sexual reproduction of commercial cultivars. Proc Int Soc Sugar Cane Technol 27:118

Oltmanns H, Frame B, Lee L-Y, Johnson S, Li B, Wang K, Gelvin SB (2010) Generation of backbone-free, low transgene copy plants by launching T-DNA from the *Agrobacterium* chromosome. Plant Physiol 152:1158–1166

Oraby H, Venkatesh B, Dale B, Ahmad R, Ransom C, Oehmke J, Sticklen M (2007) Enhanced conversion of plant biomass into glucose using transgenic rice-produced endoglucanase for cellulosic ethanol. Transgenic Res 16:739–749

Parrot DL, Anderson AJ, Carman JG (2002) *Agrobacterium* induces plant cell death in wheat (*Triticum aestivum* L.). Physiol Mol Plant Pathol 60:59–69

Patade VY, Suprasana P (2008) Radiation induced *in vitro* mutagenesis for sugarcane improvement. In: Suprasana P (ed) Special issue on biotechnology, Springer India, Sugar Tech 10:14–19

Patade VY, Suprasana P, Bapat VA (2008) Gamma irradiation of embryogenic callus cultures and *in vitro* selection for salt tolerance in sugarcane (*Saccharum officinarum* L.). Agric Sci (China) 7:101–105

Petersen K, Leah R, Knudsen S, Cameron-Mills V (2002) Matrix attachment regions (MARs) enhance transformation frequencies and reduce variance in transgene expression in barley. Plant Mol Biol 49:45–58

Petrasovits LA, Purnell MP, Nielsen LK, Brumbley SM (2007) Production of polyhydroxybutyrate in sugarcane. Plant Biotechnol J 5:162–172

Porteus MH (2009) Plant biotechnology: zinc fingers on target. Nature 459:337–338

Porteus MH, Carroll D (2005) Gene targeting using zinc finger nucleases. Nat Biotechnol 23:967–973

Potier BAM, Birch RG (2001) Sugarcane plant promoters to express heterologous nucleic acids. International Patent Publication WO/01/18211 A1

Potier BAM, Baburam C, Jacob R, Huckett BI (2008a) Stem-specific promoters from sorghum and maize for use in sugarcane. Proc S Afr Sug Technol Assoc 81:508–512

Potier BAM, Snyman SJ, Jacob R, Dheopursad D, Hucket BI (2008b) Strategies for the alleviation of promoter silencing in sugarcane. Proc S Afr Sug Technol Assoc 81:482–485

Premachandran MN (2006) Cauliflower gene in sugarcane? Curr Sci 91:750–751

Privalle LS, Wright M, Reed J, Hansen G, Dawson J, Dunder EM, Chang Y-F, Powell ML, Meghji M (2000) Phosphomannose isomerase, a novel plant selection system: mode of action and safety assessment. In: Fairbain G, Scoles A (eds) International symposium on biosafety of genetically modified organisms. University Extension Press, University of Saskatchewan, Saskatoon, Canada, pp 171–178

Putterill J (2001) Flowering in time: genes controlling photoperiodic flowering in Arabidopsis. Philos Trans R Soc Lond Biol Sci 356:1761–1767

Rajeswari S, Thirugnanakumar S, Anandan A, Krishnamurthi M (2009) Somaclonal variation in sugarcane through tissue culture and evaluation for quantitative and qualitative traits. Euphytica 168:71–80

Ralph J, Akiyama T, Kim H, Lu F, Schatz PF, Marita JM, Ralph SA, Srinivasa Reddy MS, Chen F, Dixon RA (2006) Effects of coumarate 3-hydroxylase down-regulation on lignin structure. J Biol Chem 28:8843–8853

Rathus C, Birch RG (1992) Stable transformation of callus from electroporated sugarcane protoplasts. Plant Sci 82:81–89

Rossouw D, Bosch S, Kossmann JM, Botha FC, Groenewald J-H (2007) Down-regulation of neutral invertase activity in sugarcane cell suspension cultures leads to increased sucrose accumulation. Funct Plant Biol 34:490–498

Roy PK, Kabir MH (2007) *In vitro* mass propagation of sugarcane (*Saccharum officinarum* L.) var. Isd 32 through shoot tips and folded leaves culture. Biotechnol 6:588–592

Sahrawy M, Avila C, Chueca A, Canovas FM, Lopez-Gorge J (2004) Increased sucrose level and altered nitrogen metabolism in *Arabidopsis thaliana* transgenic plants expressing antisense chloroplastic fructose-1,6-bisphosphate. J Exp Bot 55:2495–2503

Sainz MB (2009) Commercial cellulosic ethanol: the role of plant-expressed enzymes. In Vitro Cell Dev Biol Plant 45:314–329

Sakamato T, Morinaka Y, Ohnishi T, Sunohara H, Fujikota S, Ueguchi-Tanaka M, Mizutani M, Sakata K, Takatsuto S, Yoshida S, Tanaka K, Kitano H, Matsuoka M (2006) Erect leaves caused by brassinosteroid deficiency increase biomass production and grain yield in rice. Nat Biotechnol 24:105–109

Salehi H, Ransom CB, Oraby HF, Seddighi Z, Sticklen MB (2005) Delay in flowering and increase in biomass of transgenic tobacco expressing the *Arabidopsis* floral repressor gene *FLOWERING LOCUS C*. J Plant Physiol 162:711–717

Scheller J, Conrad U (2005) Plant-based material, protein and biodegradable plastic. Curr Opin Plant Biol 8:188–196

Sétamou M, Bernal JS, Legaspi JC, Mirkov TE, Legaspi BC (2002) Evaluation of lectin-expressing transgenic sugarcane against stalk borers (Lepidoptera: Pyralidae): effects on life history parameters. J Econ Entomol 95:469–477

Shiermeyer A, Shillberg S (2010) Pharmaceuticals. In: Kempken F, Jung C (eds) Genetic modification of plants. Biotechnology in agriculture and forestry, vol 64, Springer-Verlag, Berlin, pp 221–235

Shukla VK, Doyon Y, Miller JC, DeKelver RC, Moehle EA, Worden SE, Mitchell JC, Arnold NL, Gopalan S, Meng X, Choi VM, Rock JM, Wu Y-Y, Katibah GE, Zhifang G, McCaskill D, Simpson MA, Blakeslee B, Greenwalt SA, Butler HJ, Hinkley SJ, Zhang L, Rebar EJ, Gregory PD, Urnov FD (2009) Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. Nature 459:437–441

Simmons BA, Loqué D, Ralph J (2010) Advances in modifying lignin for enhanced biofuel production. Curr Opin Plant Biol 13:313–320

Snyman SJ (2004) Transformation of sugarcane. In: Curtis IS (ed) Transgenic crops of the world – essential protocols. Kluwer Academic Publishers, The Netherlands, pp 103–114

Snyman SJ, Meyer GM, Carson D, Botha FC (1996) Establishment of embryogenic callus and transient gene expression in selected sugarcane varieties. S Afr J Bot 62:151–154

Snyman SJ, Watt MP, Huckett BI, Botha FC (2000) Direct somatic embryogenesis for rapid, cost effective production of transgenic sugarcane (*Saccharum* spp. hybrids). Proc S Afr Sugar Technol Assoc 74:186–187

Snyman SJ, Meyer GM, Richards JM, Haricharan N, Ramgareeb S, Huckett BI (2006) Refining the application of direct embryogenesis in sugarcane: effect of the developmental phase of leaf disc explants and the timing of DNA transfer on transformation efficiency. Plant Cell Rep 25:1016–1023

Spiker S, Thompson WF (1996) Nuclear matrix attachment regions and transgene expression in plants. Plant Physiol 110:15–21

Sticklen M (2006) Plant genetic engineering to improve biomass characteristics for biofuels. Curr Opin Biotechnol 17:315–319

Sticklen M (2008) Plant genetic engineering for biofuel production: towards affordable cellulosic ethanol. Nat Rev Genet 9:433–443

Strauch E, Wohlleben W, Pühler A (1988) Cloning of a phosphinothricin N-acetyltransferase gene from *Streptomyces viridochromogenes* Tü494 and its expression in *Streptomyces lividans* and *Escherichia coli*. Gene 63:65–74

Taylor GO, Joyce PA, Sedl JM, Smith GR (1999) Laboratory crystallized sugar from genetically engineered sugarcane does not contain transgenic DNA in final product. Proc Austral Soc Sugar Cane Technol 21:502

Taylor LE, Dai Z, Decker SR, Brunecky R, Adney WS, Ding S-Y, Himmel ME (2008) Heterologous expression of glycosyl hydrolases *in planta*: a new departure for biofuels. Trends Biotechnol 268:413–424

Thole V, Worland B, Snape JW, Vain P (2007) The pCLEAN dual binary system for *Agrobacterium*-mediated plant transformation. Plant Physiol 145:1211–1219

Thompson CJ, Movva NR, Tizard R, Crameri R, Davies JE, Lauwereys M, Botterman J (1987) Characterization of the herbicide-resistance gene *bar* from *Streptomyces hygroscopicus*. EMBO J 6:2519–2523

Townsend JA, Wright DA, Winfrey RJ, Fu F, Maeder ML, Joung JK, Voytas DF (2009) High-frequency modification of plant genes using engineered zinc-finger nucleases. Nature 45:442–445

Trujillo LE, Sotolongo M, Menéndez C, Ochogavía ME, Coll Y, Hernández I, Borrás-Hidalgo O, Thomma BPHJ, Vera P, Hernández L (2008) SodERF3, a novel sugarcane ethylene responsive factor (ERF), enhances salt and drought tolerance when overexpressed in tobacco plants. Plant Cell Physiol 49:512–525

Ulian E (2006) Genetic manipulation of sugarcane. In: Published abstracts of the fifth molecular biology workshop of the international society of sugar cane technologists, Réduit, Mauritius, 3–4 April 2006

Umezawa T, Fujita M, Fujita Y, Yamaguchi-Shinozaki K, Shinozaki K (2006) Engineering drought tolerance in plants: discovering and tailoring genes unlock the future. Curr Opin Biotechnol 17:113–122

van der Merwe MJ, Groenewald JH, Botha FC (2003) Isolation and evaluation of a developmentally regulated sugarcane promoter. Proc S Afr Sugar Cane Technol 77:146–169

van der Merwe MJ, Groenewald JH, Stitt M, Kossmann J, Botha FC (2010) Down-regulation of pyrophosphate: D-fructose 6-phosphate 1-phosphotransferase activity in sugarcane culms enhances sucrose accumulation due to elevated hexose-phosphate levels. Planta 231:595–608

Veena JH, Doerge RW, Gelvin SB (2003) Transfer of T-DNA and Vir proteins to plant cells by *Agrobacterium tumefaciens* induces expression of host genes involved in mediating transformation and suppresses host defense gene expression. Plant J 35:219–236

Vickers JE, Grof CPL, Bonnett GD, Jackson PA, Knight DP, Roberts SE, Robinson SP (2005a) Overexpression of polyphenol oxidase in transgenic sugarcane results in darker juice and raw sugar. Crop Sci 45:354–362

Vickers JE, Grof CPL, Bonnett GD, Jackson PA, Morgan TE (2005b) Effects of tissue culture, biolistic transformation and introduction of PPO and SPS gene constructs on performance of sugarcane clones in the field. Aust J Agric Res 56:57–68

Waldron J, Reyes MEC, Hamerli D, Birch RG, Carroll BJ (2001) Tomato DNA sequences for resisting transgene silencing in sugarcane. Proc Int Soc Sug Technol 24:665–666

Wang W, Vinocur B, Altman A (2003) Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. Planta 218:1–14

Wang ML, Goldstein C, Su W, Moore PH, Albert HH (2005a) Production of biologically active GM-CSF in sugarcane: a secure biofactory. Transgenic Res 14:167–178

Wang ZZ, Zhang SZ, Yang BP, Li YR (2005b) Trehalose synthase gene transfer mediated by *Agrobacterium tumefaciens* enhances resistance to osmotic stress in sugarcane. Sugar Tech 7:49–54

Wang M, Borth W, Mangwende T, Mirkov TE, Hu J, Moore PH, Albert HH (2006) SCYLV P0: function and potential use to control transgene silencing in sugarcane. In: Published abstracts of the tropical crop biotechnology conference, Cairns, Australia, 16–19 August 2006

Wang M, Mangwende T, Borth W, Mirkov TE, HU J, Moore PH, Albert HH (2007) Constitutive expression of viral suppressors of PTGS in sugarcane. In: Published abstracts of the plant and animal genome 15th conference. San Diego, California, USA, W193, 13–17 January 2007

Warzecha H, Henning A (2010) Plastid transformation. In: Kempken F, Jung C (eds) Genetic modification of plants, Biotechnology in agriculture and forestry, vol 64. Springer-Verlag, Berlin, pp 23–37

Wei H, Albert HH, Moore PH (1999) Differential expression of sugarcane polyubiquitin genes and isolation of promoters from two highly-expressed members of the gene family. J Plant Physiol 155:513–519

Wei H, Wang M-L, Moore PH, Albert HH (2003) Comparative expression analysis of two sugarcane polyubiquitin promoters and flanking sequences in transgenic plants. J Plant Physiol 160:1241–1251

Weng LX, Deng H, Xu JL, Wang LH, Jiang Z, Zhang HB, Li Q, Zhang LH (2006) Regeneration of sugarcane elite breeding lines and engineering of stem borer resistance. Pest Manag Sci 62:178–187

Willmitzer L (1999) Plant biotechnology: output traits the second generation of plant biotechnology products is gaining momentum. Curr Opin Biotech 10:161–162

Wu L, Birch R (2005) Characterization of the highly efficient sucrose isomerase from *Pantoea dispersa* UQ68J and cloning of the sucrose isomerase gene. Appl Environ Microbiol 71:1581–1590

Wu L, Birch R (2007) Doubled sugar content in sugarcane plants modified to produce a sucrose isomer. Plant Biotechnol J 5:109–117

Xing A, Zhang Z, Sato S, Staswick P, Clemente T (2000) The use of the two T-DNA binary system to drive marker-free transgenic soybeans. In Vitro Cell Dev Biol Plant 36:456–463

Yang M, Bower R, Burow MD, Paterson AH, Mirkov TE (2003) A rapid and direct approach to identify promoters that confer high levels of gene expression in monocots. Crop Sci 43:1805–1813

Yu W, Lamb JC, Han F, Birchler JA (2006) Telomere-mediated chromosomal truncation in maize. Proc Natl Acad Sci USA 103:17331–17336

Yu W, Han F, Birchler JA (2007) Engineered minichromosomes in plants. Curr Opin Biotechnol 18:425–431

Zhang L, Xu J, Birch RG (1999) Engineered detoxification confers resistance against a pathogenic bacterium. Nat Biotechnol 17:1021–1024

Zhang D, Li X, Zhang LH (2002) Isomaltulose synthase from *Klebsiella* sp. strain LX3: gene cloning and characterization and engineering of thermostability. Appl Environ Microbiol 68:2676–2682

Zhang D, Li N, Swaminathan K, Zhang LH (2003) A motif rich in charged residues determines product specificity in isomaltulose synthase. FEBS Lett 534:151–155

Zhang JZ, Creelman RA, Zhu JK (2004) From laboratory to field. Using information from Arabidopsis to engineer salt, cold, and drought tolerance in crops. Plant Physiol 135:615–621

Zhang SZ, Yang BP, Feng CL, Chen RK, Luo JP, Cai WW, Liu FH (2006) Expression of the *Grifola frondosa* trehalose synthase gene and improvement of drought-tolerance in sugarcane (*Saccharum officinarum* L). J Integr Plant Biol 48:453–459

Zhang X, Du P, Lu L, Xiao Q, Wang W, Cao X, Ren B, Wei C, Li Y (2008) Contrasting effects of Hc-Pro and 2b viral suppressors from *Sugarcane mosaic virus* and *Tomato as-permy cucumovirus* on the accumulation of siRNAs. Virology 374:351–360

Zhangsun D, Luo S, Chen R, Tang K (2007) Improved *Agrobacterium*-mediated genetic transformation of GNA transgenic sugarcane. Biologia 62:386–393

Zhao Y, Qian Q, Wang HZ, Huang DN (2007) Co-transformation of gene expression cassettes via particle bombardment to generate safe transgenic plant without any unwanted DNA. In Vitro Cell Dev Biol Plant 43:328–334

Chapter 12
Genetic Engineering of *Miscanthus*

Dean Engler and Katrin Jakob

Abstract This chapter describes the advantages and present limitations of developing transgenic *Miscanthus* genotypes with improved characteristics for the emerging biomass and biofuels industries. An efficient method for transformation of *Miscanthus* developed in the laboratories of the authors is presented. Traits of value to the biomass/biofuels industry and strategies for how they could be improved by insertion of a transgene(s) are described, including herbicide resistance, biotic and abiotic stress resistance, and biomass composition improvements. In addition, we describe strategies for transgenically controlling traits leading to improvements in yield parameters such as plant height, tiller number, branching patterns, and time of flowering. Methods to integrate transgenic genotypes into a breeding program are discussed.

Keywords Biomass • Cellulose • Cellulosic biofuels • Flowering control • Geneticin • Lignin • *Miscanthus* • *Giganteus* • Plant stress • Transgenic

1 Background

Miscanthus has recently emerged as a potential crop of significance in temperate regions including the USA (Heaton et al. 2004a, b) which could be grown as an alternative to coal for energy generation, and as a feedstock for lignocellulosic ethanol production. *Miscanthus* has been cultivated as an energy plant in Europe for over a decade, and it has been demonstrated to be particularly suitable as a temperate biomass feedstock plant because of its perenniality, highly efficient photosynthetic

D. Engler (✉) • K. Jakob
Mendel Biotechnology, Inc, 3935 Point Eden Way, Hayward, CA 94545-3720, USA
e-mail: dengler@mendelbio.com

A.H. Paterson (ed.), *Genomics of the Saccharinae*, Plant Genetics and Genomics: Crops and Models 11, DOI 10.1007/978-1-4419-5947-8_12,

capacity, low need for agronomic inputs, adaptation for growth on marginal lands, and very high potential yield (Clifton-Brown et al. 2001; Christian et al. 2008). *Miscanthus* yields significantly exceed those of Switchgrass, another candidate biofuel crop, in many environments (Heaton et al. 2004a, b). Unlike *Miscanthus*, however, switchgrass is not in the *Saccharinae* clade.

Genetic engineering of potential biofuels crops such as species of *Miscanthus* is an attractive possibility. *Miscanthus* × *giganteus* is a prime candidate as a source of biomass, but is triploid and totally sterile, and therefore improvement by conventional intercrossing among genotypes to pyramid favorable genes is not possible. Fertile *Miscanthus sinensis*, one of the progenitors of *M.* × *giganteus* has received some breeding attention, but to date largely for the development of ornamental varieties for gardens and not for the biomass/biofuels industries. Other species such as *Miscanthus sacchariflorus* have received even less attention from scientists and horticulturalists.

There are many characteristics of critical importance to biofuels crops that can likely be improved by the presence of one or more transgenes. These characteristics include improved yield, resistance to stresses (e.g., disease, temperature, and drought), resistance to herbicides, flowering control, and compositional changes that may improve caloric content or convertibility to liquid fuels (e.g., lignin modification). Candidate genes are known which have the potential to affect these characteristics, and future transgenic *Miscanthus* varieties with such improvements are likely to be produced.

The first US field with transgenic *Miscanthus* for improved nitrogen efficiency has been approved in 2011(Status of permits, notifications, petitions, USDA-APHIS 2012) while no transgenic *Miscanthus* trial has been conducted in Europe yet (Environmental releases of GMOs, European Union 2012). Multiple field trial permits have been granted for field releases of transgenic Switchgrass in the USA, targeting multiple traits including drought tolerance, nitrogen use efficiency, male sterility, herbicide tolerance, and lignin modification (Status of permits, notifications, and petitions, USDA-APHIS 2012).

Many graminaceous monocot transformation methods rely on regenerable embryogenic callus (ecallus) as the tissue targeted for DNA delivery (Wang and Ge 2006). In these methods, ecallus is generally initiated from somatic explants such as immature spikelets, mature or immature embryos, or seedling tissues. The resultant ecallus is cultured for many weeks or months until a sufficient amount is present to allow an attempt to deliver exogenous DNA. It is critically important to develop methods to foster the continual ability of the ecallus to produce regenerated plants during this "bulking" process, the DNA delivery process, and finally the process of selection for transgenic derivatives of the original ecallus cells.

Several delivery methods are available for transferring DNA to ecallus cells. A commonly used method is particle bombardment (e.g., Frame et al. 2000). A second method which has been used on ecallus initiated from many diverse species is cocultivation with *Agrobacterium* species (most commonly *A. tumefaciens*). The long list of plant species which have been transformed with this method includes many important dicots such as cotton (Khan et al. 2010), grapes (Lopez-Perez et al. 2009),

several citrus species (e.g., Duan et al. 2007), roses (Korban, et al. 2006), and soybeans (Trick and Finer 1998). In addition, many monocot species have been transformed by cocultivation of ecallus, including bananas (Ghosh et al. 2009), elite lines of maize (Yang et al. 2006), onions (Aswath et al. 2006), switchgrass (Xi et al. 2009), rice (Kumria et al. 2001), sugarcane (Wang et al. 2005), and several species of turfgrasses (Longo et al. 2006).

A method for *M. sacchariflorus* transformation, which is based on particle bombardment of ecallus, has been published (Zili et al. 2004). A patent application which describes a method for *Miscanthus* transformation which was developed in our laboratory by cocultivation of ecallus with *Agrobacterium tumefaciens* has also been published (Engler and Chen 2009). Both of these publications, in agreement with an earlier paper (Holme and Petersen 1996), designate immature terminal spikelets as the preferred explant for production of regenerable ecallus. Zili et al. (2004) bombarded the ecallus with DNA-coated gold particles in a PDS-1000/He particle gun followed by selection for transformation events on 10 mg/l Hygromycin. In our method (Engler and Chen 2009) ecallus is cocultivated with the *A. tumefaciens* strain GV3101 (pMP90) followed by selection for transformation events on 100 ppm G418 (Geneticin). In both publications, transformed ecallus was produced, successfully selected, and regeneration of the transformed embryogenic calli resulted in production of transgenic plants.

The use of particle bombardment for DNA delivery into plant cells has the disadvantage that many copies of the transferred sequence are routinely integrated into the targeted genome. These integrated copies are often rearranged and mutated. Further, the integrated sequences are often unstable (Casas et al. 1993). This can lead to low expression levels and/or instability of the trait of interest, and also may lead to difficulty in gaining acceptance of the transgene locus with regulatory agencies.

Miscanthus is susceptible to many of the antibiotics and herbicides that are often used for selection of transgenic tissues following a transformation protocol. Hygromycin, Basta, and G418 were shown to be appropriate selective agents when used with their corresponding resistance genes (Zili et al. 2004; Engler and Chen 2009). Unlike other graminaceous monocots (Wilmink and Dons 1993), even at high concentrations, Kanamycin was not an effective selection agent in *Miscanthus*, and therefore, employment of the NPTII gene will generally require the use of an alternative antibiotic such as Geneticin or Paromomycin (Engler and Chen 2009).

2 *Miscanthus* Transformation

2.1 Miscanthus Transformation Protocol

The steps necessary for production of transformed *Miscanthus* genotypes are:

1. Production of regenerable ecallus.
2. Culture of regenerable ecallus.

3. DNA delivery into the cells of regenerable ecallus.
4. Culture and selection of transformed *Miscanthus* ecallus cells.
5. Regeneration of whole transgenic plants from the transformed ecallus cells.

2.1.1 Production of Regenerable Embryogenic Callus from Seeds

Miscanthus seeds inoculated onto ecallus induction media produce ecallus quite readily, and at an acceptable frequency (depending on genotype, 5–10% of seeds may produce ecallus). Use of seed-derived ecallus may be acceptable for investigations into transformation methods or for application of other academic pursuits, but it would have little value for the production of a commercially significant transgenic product. This is due to the genetic heterogeneity of seeds produced as a result of crossing or the rare self pollination of this self incompatible and therefore largely outcrossing species. All plants, including transgenic plants regenerated from such ecallus, have unknown agronomic value and would be unlikely to form the basis of a new variety, either directly (by cloning) or indirectly as one parent of a hybrid.

If seeds are to be used as the explant for production of ecallus, they must first be sterilized. Sterile seeds can be produced by immersing *M. sinensis* seeds in a 20% bleach solution to which 0.1% Triton-X100 has been added as a wetting agent for 20 min, followed by five rinses in sterilized distilled water. A commercially available source of *M. sinensis* is "Jelitto perennial seeds" located in Schwarmstedt, Germany. We have found that seeds of the Jelitto variety "Pure Seed" produce ecallus at a high frequency. The sterilized seeds are then plated onto *Miscanthus* ecallus induction and growth medium (MECG; Engler and Chen 2009) in sealed petri dish plates (100 mm × 25 mm). The plates are incubated in continuous darkness at 29°C for 6 weeks, after which high quality ecallus can be visually selected with the aid of a dissecting microscope.

2.1.2 Production of Embryogenic Callus from Immature Spikelets

As discussed above, seed-derived ecallus is easy to produce, but has less practical value than ecallus derived from somatic tissues. As demonstrated by Holme and Petersen (1996) and also by Glowacka and Jezowski (2010), ecallus can be produced from the somatic tissues present in immature spikelets of *Miscanthus* tillers which are just beginning to flower. We have found the Jelitto variety "Late Hybrid" suitable for production of ecallus from immature spikelets. Tillers which have 5–6 fully open leaves are selected, and the top internode is detached and sterilized by immersion in a bleach/Triton X100 solution, as described above, for 5 min followed by five rinses in sterilized distilled water. The immature spikelets are removed from the top internode, cut into segments, and plated on MECG media in sealed petri dish plates (100 × 25 mm). These plates should be incubated at 29°C for 6 weeks in continuous darkness and examined as above for the presence of ecallus.

2.1.3 Culture of Regenerable Embryogenic Callus

ecallus, whether seed- or spikelet-derived, is incubated on MECG in continuous darkness at 29°C. The ecallus should be subcultured to fresh MECG approximately every 3–4 weeks. Many cell types form during these incubation periods, and it is critical to select the type of cells that have the capacity to regenerate whole plants for subculturing. Highly skilled tissue culturists may be able to recognize the callus morphology types which maintain regenerability. These morphological types have been described as type I and type II calli, and these types of calli have been described in multiple publications with multiple graminaceous species (Merkele et al. 1995). Some examples of these publications for corn are Lu et al. (1982); Songstad et al. (1992); and Welter et al. (1995). An example in rice is Chen et al. (1985). Examples of ecallus types also exist for species closely related to *Miscanthus*, e.g., sorghum (Jeoung et al. 2002) and sugarcane (Guiderdoni and Demarly 1988).

We have found that there is a simple way to ensure that *Miscanthus* callus is continually regenerable which does not rely on the identification of a specific type of callus morphology. Prior to each subculture, the plates containing the ecallus are exposed to white light provided by cool white florescent tubes (70 μmol/m^2/s) for 3–7 days. During this exposure to light, some of the ecallus will turn green, and we have found that it is this green callus that is regenerable and therefore should be preferentially selected for transfer. The callus which does not turn green after exposure to light is discarded. Using the capacity to turn green, we have maintained regenerable *Miscanthus* ecallus for greater than 2 years.

2.1.4 DNA Delivery into the Cells of Regenerable Embryogenic Callus

Transformation of *Miscanthus* ecallus can be accomplished by biolistic DNA delivery (Zili et al. 2004) or by cocultivation with *A. tumefaciens* (Engler and Chen 2009). Although intellectual property issues may lead researchers to one or the other of these methods, the preferred method would be the use of *A. tumefaciens* for the reasons discussed above. Transformation of ecallus is initiated by infection and cocultivation with *A. tumefaciens*. Several strains of *A. tumefaciens* are effective for DNA delivery. We have had success using strains GV3101 (pMP90RK) in combination with OriV-containing binary plasmids, and also with other *Agrobacterium* strains such as ABI and EHA105 when using non-OriV-based binaries. For a discussion of *Agrobacterium* strains and vector systems useful for plant transformation, see the review paper by Hellens and Mullineaux (2000). It is useful to include a β-glucuronidase (GUS; Jefferson 1989) gene as a reporter, and neomycin phosphotransferase II (NPTII) gene conferring G418 (Geneticin) resistance as a selectable marker within the T-DNA.

The *Agrobacteria* containing the chosen construct are grown in 50 ml LB liquid *Agrobacterium* growth medium (plus antibiotics as appropriate for the Agro strain being used) in a 250 ml Erlenmeyer flask at 250 rpm at 28°C overnight. Young *M. sinensis*

"Pure Seed" ecallus grown on MECG at a size of about 3 mm in diameter are selected under a dissection microscope. The stage of growth of the ecallus used for cocultivation was immediately following exposure to light. The selection is accomplished using ecallus morphological patterns combined with chlorophyll biosynthesis as a selection marker for regenerable ecallus. The *Agrobacteria* grown overnight are diluted to OD 0.6 with MECG liquid medium. The selected ecallus are infected by immersion in the *Agrobacterium* liquid for 5–10 min in a petri dish. The *Agrobacterium* liquid is removed by sterile pipette. The ecallus is then transferred on to "*Agrobacterium* and ecallus cocultivation medium" (MECC; Engler and Chen 2009) in petri dish plates for cocultivation. The plates for cocultivation are incubated in the dark at 25°C for 5 days.

2.1.5 Culture and Selection of Transformed *Miscanthus* Embryogenic Callus Cells

After 5 days of cocultivation, embryogenic calli are transferred onto "Transgenic ecallus selection medium" (MECS). MECS consists of MECG plus 100 ppm G418 (Geneticin) for NPTII gene selection, and 150 ppm Timentin to eliminate agrobacteria. Both of these ingredients are filter sterilized prior to adding to MECG. The plates of cocultivated embryogenic calli are incubated in continuous darkness at 29°C. After 3 weeks, the embryogenic calli are transferred to fresh MECS under the same growth conditions. After another 3 weeks, the newly formed embryogenic calli can be assayed for the presence of the GUS enzyme which is indicative of a stably integrated GUS gene.

2.1.6 Regeneration of Whole Transgenic Plants from the Transformed Embryogenic Callus Cells

After a total of 6 weeks selection on MECS in the dark, the plates can be moved to continuous white light provided by cool white florescent tubes (70 μmol/m^2/s) for 2 weeks. The green plants that form are transferred to hormone-free MS media supplemented with 150 mg/l Timentin for further growth and rooting. It is possible to include G418 in this media, and escapes are virtually never seen when this is done. Leaf tips of regenerated whole plants can be assayed for the presence of GUS at this time. Transformed plant leaf tips will turn blue in color when exposed to the GUS stain solution within a few hours. Whole, rooted *Miscanthus* plants are planted in pots filled with autoclaved "sunshine" soil mixture (Western Farm Supply) which is covered to prevent the tender plants from wilting. When the plants are actively growing, they can be "hardened off" by gradual reduction of humidity, and finally they can be transferred into a greenhouse for growth and seed production.

2.2 Results of Miscanthus Transformation

The above protocol has been applied to *M. sinensis* multiple times at Mendel Biotechnology, and transformed plants are routinely produced. We have found that the overall transformation efficiency (defined as the number of transgenic plants produced expressed as a percentage of the number of ecallus clusters cocultivated) is approximately 4.0%. Nontransgenic embryogenic calli which have escaped the selection are rarely seen. No shoots will form if these rare escape calli are transferred to regeneration media containing the selective agent. No escape plants have ever been produced as a result of regenerating shoots on media containing the selection agent. In total, 25 independently transformed GUS expressing plants have been produced to date.

3 Traits of Interest for *Miscanthus*

Efficient plant breeding programs will be required to develop the undomesticated *Miscanthus* as a high yielding and profitable bioenergy crop that can be grown at different geographies. Within these breeding programs, the application of transgene technology constitutes an effective tool to significantly accelerate the development of *Miscanthus*. Moreover, genetic modification of *Miscanthus* can build on platforms that have been developed and successfully deployed for other crops such as corn, soybean and cotton. In 2011, 160 million hectares were planted with biotech crops, and of these, crops grown on about 42 million hectares contained stacked traits (ISAAA 2012). The latest generation of biotech plants released contains eight genes conferring multiple insect as well as broad-spectrum weed control (www.monsanto.com). One of the major hurdles in transgene technology is transformation and transformation efficiency. Since transformation has been proven feasible and effective (see above), genetic engineering of *Miscanthus* could be an integral part of *Miscanthus* genetic improvement.

Knowledge of the genetics underlying traits of interest is essential to enable genetic modification of *Miscanthus* for improved plant performance. *Miscanthus* is a new crop with almost no genetic information available. However, the importance of the crop as biofuel feedstock was recognized early on by the plant community and whole genome sequencing of the standard clone *M.* × *giganteus* is underway (www.jgi.doe.gov). A first survey of the *M.* × *giganteus* transcriptome revealed significant sequence similarity to sorghum and confirmed the presence of many repetitive sequences characteristic of grasses, complicating sequence assembly (Swaminathan et al. 2010). Sorghum and sugarcane are the crops most closely related to *Miscanthus* and provide a wealth of genetic information, including a fully sequenced genome for sorghum (Paterson et al. 2009) and simple sequence repeat (SSR) resources for both crops. Genome synteny has been demonstrated for genes and quanititive trait loci (QTLs) in rice, sorghum and other plants and it is anticipated

that genes of interest in *Miscanthus* can be mapped in corresponding genome regions of sorghum once genetic maps of *Miscanthus* become available. Among the choice of molecular markers that can be applied for gene discovery and molecular marker assisted breeding, SSRs are particular suitable for crops with little sequence availability, as sequences in the flanking regions of SSRs are often conserved within and between species. A first set of *Miscanthus* SSRs has been identified based on sugarcane SSRs (De Cesare et al. 2010) and by sequencing SSR enriched *Miscanthus* DNA libraries (Hung et al. 2009). Further sequencing of *Miscanthus* will eventually enable the application of single nucleotide polymorphism (SNP) markers for gene discovery, trait mapping and subsequently the transfer of these genes into *Miscanthus*. In fact, while this manuscript was in review, genetic maps have been published for *M. sinensis* and *sacchariflorus* based on SSRs (Kim et al. 2012). In addition, 19 linkage groups could be identified for *Miscanthus sinensis* by applying genotyping by sequencing (Ma et al. 2012). The successful alignment of *Miscanthus* linkage groups to sorghum chromosomes revealed evidence for a whole genome duplication event in *Miscanthus* and further supports the application of sorghum genomic resources for accelerating the improvement of *Miscanthus*.

There are common traits of agronomic interest for row and biomass crops, such as yield and biotic and abiotic stress resistance, for which transgenic improvements have been provided. Insect and herbicide resistance are among these traits and could be readily applied in *Miscanthus*. On the other hand, there are particular biofuel traits that are quite different from traits of interest in conventional row crops. Cell wall properties like lignin, cellulose, and hemicellulose content and characteristic are most crucial for lignocellulosic biomass conversion. Although modification of the lignin pathway has long been recognized for improving digestibility in animal feedstock and pulping, genetic research in this area has been taken to new heights with regard to biofuel production. Knowledge of cellulose and in particular hemicellulose synthesis and degradation is still rather limited and implementing research results for *Miscanthus* improvement might take some time.

3.1 Application of Insect, Disease and Herbicide Resistance Available on the Market

Genetically engineered insect and herbicide resistance have been successfully incorporated into row crops, and these have been grown worldwide for many years (Dill et al. 2008; Padgette et al. 1995; Nida et al. 1996; CERA 2012). Insect resistance against corn borer and corn root worm in corn has been achieved by incorporating Cry-proteins from microbes, mainly *Bacillus thuringiensis*. *Cry1Ab*, for instance, provides resistance to the European corn borer (*Ostrinia nubilalis*), Southwestern corn borer (*Diatraea grandiosella*) and the Mediterranean stem borer (*Sesamia nonagrioides*) (Archer et al. 2001; Castro et al. 2004) and *Cry3Bb1* and the binary *Bt* toxin *Cry34*/*Cry35* protect corn from the Western corn rootworm (*Diabrotica virgifera virgifera*) (Ellis et al. 2002). The *Cry1F* protein controls the European and

Southwestern corn borer, but also protects against other lepidopteran pests including sugarcane borer (*Diatraea saccharalis*) fall armyworm (*Spodoptera frugiperda*), black cutworm (*Agrostis ipsilon*), and Western bean cutworm (*Richia albicosta*) (Catangui and Berg 2006; Eichenseer et al. 2008; Siebert et al. 2008). The latest multi gene product available on the market is based on *Cry1A*, *Cry2A*, *Cry3Bb1*, *Cry1F* and *Cry34* and provides resistance against a combination of insects like the European corn borer, the Southwestern corn borer, the Northern corn rootworm, the corn ear worm, fall armyworm, Western bean and the black cutworm (Greenberg and Adamczyk 2007; Stewart et al. 2001; Willrich et al. 2005).

Insect resistance would similarly be important in *Miscanthus*. Although no serious pests and diseases have been reported in *Miscanthus* in the USA or EU yet, some of the above mentioned pests and diseases have already been observed on *Miscanthus*. *Fusarium miscanthii* has been identified from *Miscanthus* straw in Denmark (Gams et al. 1999). The Asiatic pink stem borer S*esamia inferens* was found on *Miscanthus* (Zhou and Chen 1985) as well as wireworms (*Elateridae*) and corn borer which have been detected on *Miscanthus* hybrids in field trials (F. Zhou, K. Jakob, unpublished). Oviposition of Western corn rootworm on *Miscanthus* have been observed in the greenhouse (Spencer and Raghu 2009) presenting the possibility of *Miscanthus* being a reservoir or refuge for this corn pest. If these pests become a significant problem in *Miscanthus*, the readily available multistacked insect and disease resistance for corn should be transferable to *Miscanthus*.

A significant concern for successful establishment of new *Miscanthus* plantations is weed control in the planting year. Establishment of currently available biomass *Miscanthus* is slow compared to sorghum, corn and other annual crops, and young *Miscanthus* plantlets can easily be overgrown by weeds causing substantial plant loss. Presently, this would apply to all forms of establishment, i.e., regardless of whether *Miscanthus* is grown from rhizomes, plantlets/plugs or seeds. In a perennial plantation, missing plants would potentiate biomass loss over the years if the neighboring plants are not able to fill in the space with tillers accordingly; hence herbicide control in the first year would be desirable. However, once *Miscanthus* has been established weed control is no longer critical. Commercially available genetically engineered herbicide resistance to glyphosate, known as Round-up, conferred by CP4 EPSPS (Kishore and Shah 1988); to glufosinate (Hoerlein 1994), known as Liberty and Basta; and to sulfonylurea herbicides such as Glean and Oust provided by target site mutations in the ALS gene (Lee et al. 1988) are all available. All are broad spectrum herbicides operating systemically in the case of glyphosate and sulfonylureas, and on contact in the case of glufosinate. Although desirable for effective *Miscanthus* establishment in the field, genetically engineered herbicide resistance in *Miscanthus* has to be approached carefully as it could compromise the ability to remove perennial *Miscanthus* in a crop rotation or when it is growing where it is not desired.

Long established *Miscanthus* plantations can be removed by plowing and subsequent herbicide applications. As an example, glyphosate resistant *Miscanthus* could not be removed via Round-up application which is presently common practice. Use of alternate herbicides is possible, but this would lead to additional expense. An option to overcome this hurdle might be to express the transgene via a promoter

which would activate the gene only in an early developmental stage of *Miscanthus* in the establishment year. In addition, transfer of the transgene function via seeds or vegetative plant parts needs to be prevented.

Commercially available transgenic plants with disease resistance presently exist only for virus resistance in four crops, potato, plum, squash and papaya. These are all based on the overexpression of genes producing the virus coat protein to provide resistance against the virus. The CZW-3 squash exhibits resistance to cucumber mosaic virus (CMV), zucchini yellow mosaic potyvirus (ZYMV) and watermelon mosaic virus (WMV2) through a process that is related to viral cross-protection (Register and Nelson 1992) thereby increasing yield up to 50% (CERA 2012). In papaya, resistance to papaya ringspot virus (PRSV) is also achieved by encoding the PRSV coat protein in the transgenic plant (Gonsalves 1998). Resistance to plum pox virus (PPV) in plum is available but has not been introduced into the USA (CERA 2012). Barley yellow dwarf virus has been detected in *Miscanthus* (Christian et al. 1994) as well as sugarcane mosaic virus in *M.* × *giganteus* (Agindotan et al. 2010) and there is a chance that other viruses including those known to exist in sugarcane and grain crops could develop as pathogens on *Miscanthus*.

3.2 Modification of Lignin, Cellulose, and Hemicelluloses to Overcome Recalcitrance

Rapid advancement of a cellulosic biofuels industry has largely been hindered by technical challenges in processing biomass into fuel. Chemical conversion of lignocellulosic feedstock involves lignin degradation followed by saccharification of the polysaccharides into C5 and C6 monomers and fermentation of these sugars into fuel. To enable access to the sugar molecules, lignin has to be degraded which is hampered by the strong cross linking to cellulose and hemicellulose. Current technology pretreats biomass with strong acid, followed by enzyme application, resulting in hydrolysis of the polysaccharides and fermentation. Processing of lignocellulosic plants is about twice as expensive as conversion of corn starch to sugar (Lynd et al. 2008), which has caused a lag in the transformation from sugar/starch ethanol to lignocellulosic ethanol production. In addition to the ongoing technical improvement of processing, biotechnology could play a major role in addressing these obstacles and significantly impact the advancement of lignocellulosic biofuels. The analysis of different processing technologies identified the biggest cost saving potential in the first processing step, the conversion of biomass into sugars which involves the breakdown of lignin, ideally without pretreatment, and cellulose hydrolysis (Lynd et al. 2008).

The “ideal” lignocellulosic biomass for ethanol production would produce cell walls with low lignin, high cellulose and hemicellulose content, and easy access to these plant cell wall components for hydrolysis. Alternatively, thermal conversion of lignocellulosic plants for the bioenergy market (e.g., by burning) would require maximum lignin content which translates into a higher energy value per unit biomass. Further, plant endogenous enzymes could activate hydrolysis of these

polysaccharides at a defined time and also provide and activate enzymes for the conversion of sugars into fuel. While each approach has been taken separately at the research level, combining all three at production scale will remain a challenge. However, stepwise progress may significantly reduce costs for each technical production process. Lignin, cellulose, and hemicellulose are the main polymers incorporated into plant cell walls, and their composition and ratio varies widely between cell wall types within plants, plant species, and environments. Cellulose and hemicelluloses comprise the sugar source for fermentation into fuel. Lignin is bound to hemicelluloses and cellulose filling in the space between the polysaccharides, thereby stabilizing the plant cell wall, enabling water transport and protecting the plant against pathogen attacks. Although pectin is a large part of dicotyledonous primary cell walls, it is found only in smaller amounts in secondary cell walls of grasses.

Biotechnical approaches to optimize biomass convertibility could be initiated at three stages in the life of the grass cell wall. Modifying the synthesis of the primary substrates required for the formation of cellulose, hemicelluloses, and lignin could increase carbon levels for conversion. Diverting the synthesis of the polymers could reduce lignin and change the ratio in favor of cellulose and hemicelluloses. Modifying polymer synthesis could reduce crystallinity and reduce the strength of the bonds between the molecules in lignin and cellulose to make it less resistant to degradation, enable hydrolysis, and provide enzymes that transform polymers into simple sugars for enhanced saccharification. This latter step might obviate the need for adding external enzymes and possibly provide enzymes promoting the fermentation of sugars into the final product.

Cellulose, hemicelluloses, and lignin synthesis and incorporation of the molecules into the cell wall is facilitated by highly complex pathways. Historically, the lignin pathway has been studied more intensively for feedstock improvement and pulping, and genes of the lignin pathway have been completely identified over the last three decades (Hisano et al. 2009; Simmons et al. 2010). Cellulose and hemicellulose pathways, as well as the transcriptional regulation of all three pathways, remain to be decoded. Studies of the cellulose pathway seem much more challenging due to lethal or infertile cell wall mutants. With the identification of the cellulose synthase (CESA) complex and CESA-like genes synthesizing cellulose, cellulose modification is now becoming feasible. The hemicellulose pathway discovery is still in its infancy and even more complex because hemicellulose consists of not only one but several different polymers. For the genetic modification of *Miscanthus*, lignin reduction would be highly desirable to enable the elimination of costly pretreatments for cellulosic ethanol production. Gradually, findings from cellulose and hemicellulose research could be incorporated into biotechnical approaches to improve *Miscanthus* for processing. The technology for processing cellulosic biomass into ethanol is still under development and standardized requirements for lignin, cellulose, and hemicellulose content and properties for optimal processing are not available yet. Likewise, on the plant site, there are more studies needed to learn how different levels of the cell wall components affect plant growth. For ethanol production, the goal is to decrease lignin and increase cellulose and hemicellulose content, and decrease recalcitrance for better degradation.

3.2.1 Lignin

The average lignin content of dry, chopped and milled *Miscanthus* straw varied between 9.9 and 18% for *M.* × *giganteus*, 12.5 and 19.7% for *M. sacchariflorus* and 10.4 and 16.7% for *M. sinensis* in a 3 year field experiment (Pude 2005). Cellulose content ranged from 37.9 to 55.6% for *M.* × *giganteus*, 37.7–48.3% *M. sacchariflorus* and 34.5–46.4% for *M. sinensis* (Pude 2005). Higher lignin and cellulose content in *M.* × *giganteus* and *M. sacchariflorus* compared to *M. sinensis* was also observed in a 15 location trial in Europe, however the latter was not significant (Hodgson et al. 2010). In addition, environment affected cellulose and hemicellulose content but no such effect was observed for lignin (Hodgson et al. 2010).

Lignin provides stability, enables water transport (Coleman et al. 2008) and is involved in biotic and abiotic stress responses (Ali et al. 2006; Bhuiyan et al. 2009; Menden et al. 2007; Taheri and Tarighi 2010). Lignin is deposited primarily in secondary cell walls and only smaller amounts are found in primary cell walls. Two main strategies can be employed for reducing lignin recalcitrance. Overall lignin content can be reduced and the ratio between two of the main lignin molecules, syringyl (S) and guaiacyl (G) lignin, can be modified in favor of the more easily digestible S-lignin. Lignin is produced in the monolignol pathway, and derived from three monolignols, *p*-coumaryl, coniferyl and sinapyl alcohols. Phenylalanine is the primary substrate for lignin biosynthesis, which is transformed by ten major genes in the monolignol pathway (Hisano et al. 2009). Orthologs of lignin pathway genes have been described in poplar and Arabidopsis (Xu et al. 2009), and comparative genome analysis identified monolignol pathway genes in rice, sorghum (Xu et al. 2009) and corn (Barrière et al. 2009; Penning et al. 2009). Overall, genes from eight (Penning et al. 2009) or ten (Xu et al. 2009) gene families encode the enzymes in the monolignol synthesis in Arabidopsis, rice and sorghum. An expansion of gene families in the lignin pathway was observed after the evolutionary split between monocotyledonous and dicotyledonous plants, except for the C4H (cinnamate 4-hydroxylase) gene family. This indicates that studies in dicotyledonous model plants might not be directly applicable in *Miscanthus*. High synteny was found between rice and sorghum lignin pathway genes (Xu et al. 2009). Transcriptional profiling of cell wall genes under different stress conditions comparing *Arabidopsis thaliana* and corn also demonstrated substantial differences in gene expression between monocotyledonous and dicotyledonous plants (Penning et al. 2009).

Up and down regulation of genes in the lignin pathway has successfully been demonstrated to reduce lignin content and modify its characteristics, thereby increasing digestibility in feedstock plants and pulping efficiency in woody plants (Huntley et al. 2003). Expression studies demonstrated high linkage between increased digestibility and underexpression of lignin pathway genes (Barrière et al. 2009). Downregulation of enzymes acting early in the pathway, like C4H (cinnamate 4-hydroxylase), 4CL (4-coumarate-coenzyme A ligase), HCT (hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase) and C3H (4-coumarate 3-hydroxylase) decreased lignin content, whereas F5H (ferulate

5-hydroxlase), located downstream of these genes, affected lignin composition. In particular, the ratio of p-hydroxyphenyl (H), guaiacyl (G) and syringyl (S) lignin units was altered (Sibout et al. 2002). F5H overexpression in poplar increased S-lignin to 97.5% (Stewart et al. 2009) and downregulating F5H in *A. thaliana* reduced S-lignin to almost zero (Meyer et al. 1998). F5H overexpression increased pulping efficiency due to a significantly higher S/G ratio, while no change in lignin content was observed (Huntley et al. 2003). The F5H gene was found to be distinct between monocotyledonous and dicotyledonous plants (Xu et al. 2009). The identification of an endogenous F5H gene in *Miscanthus* and overexpression of this gene could be a first approach to increase S-lignin and lignin degradation in *Miscanthus*. Similarly, overexpression of the *A. thaliana* or poplar F5H in *Miscanthus* could drive increasing S-lignin.

Transgenic aspen with reduced 4CL expression contained up to 45% less lignin and 15% more cellulose, and, in addition, showed enhanced growth (Hu et al. 1999). Downregulation of C3H by RNAi in poplar reduced total lignin content, increased glucose and xylose content (Coleman et al. 2008), and caused a decrease of the guaiacyl (G) monomer, but the S-lignin content was unaffected. In contrast, both S- and G-lignin were reduced in a similar C3H antisense mutant in *Medicago* (Reddy et al. 2005) and *Arabidopsis* (Franke et al. 2002). However, reduced growth rates were observed for the transgenic plants (Coleman et al. 2008). Antisense downregulation of a lignin specific peroxidase, which in turn downregulates lignin during polymerization, produced 20% less lignin in transgenic tobacco leading to a threefold increase of saccharification (Kavousi et al. 2010). The plant also produced less chlorophyll, a smaller number of stomata but overall produced a normal plant phenotype (Kavousi et al. 2010). In alfalfa, downregulation of each of six genes of the lignin pathway revealed highest lignin reduction when C4H, C3H and HCT were downregulated (Chen and Dixon 2007). Lowest lignin levels were directly correlated with an increase in sugar release (Chen and Dixon 2007), which was even more enhanced after stem pretreatment. Interestinlgy, the S/G ratio was not correlated with sugar release in HCT and C3H lines.

Spontaneous or chemically induced low lignin mutants have been identified in corn (Haney et al. 2008; Jorgenson 1931; Kuc and Nelson 1964; Neuffer et al. 1997), sorghum (Porter et al. 1978; Saballos et al. 2008; Xin et al. 2009) and pearl millet (Gupta 1995; Cherney et al. 1988). Low lignin mutants are called brown midrib (*bmr*) due to the phenotypic brown coloration of the leaf midrib and some of the mutations have been localized in the COMT (caffeic acid *O*-methyl transferase) and CAD (cinnamyl alcohol dehydrogenase) genes of the lignin pathway (Halpin et al. 1998; Sattler et al. 2010; Vignols et al. 1995). Interestingly, no brown midrib has been found in C3 grasses, switchgrass, or sugarcane, and it has been suggested that this might be due to redundancy in the more complex and polyploid genomes of these plants compared to corn, sorghum and pearl millet (Sattler et al. 2010). Brown midrib mutants have decreased lignin thereby increasing digestibility (Kuc and Nelson 1964; Lechtenberg et al. 1972). Although it has been shown that *bmr* mutations can reduce grain and stover yield and increase susceptibility to pests and diseases (Pedersen et al. 2005), these obstacles are expected to be overcome by hybrid breeding (Sattler et al. 2010). The search for naturally occurring brown

midrib mutants in *Miscanthus* germplasm as well as obtaining chemically induced mutations as demonstrated in sorghum (Xin et al. 2009) are also promising options for establishing low lignin mutants in *Miscanthus*.

3.2.2 Cellulose

Cellulose is comprised of a chain of glucose molecules (beta-1-4 glucan, Somerville 2006) and different chains are tightly linked by hydrogen bonds forming exceptionally strong microfibrils. Similarly to lignin, this crystalline structure is highly recalcitrant to degradation. The identification of genes and enzymes involved in cellulose synthesis has been complicated by lethality and reduced fertility of cell wall mutants, redundancy in the cellulose synthesizing enzymes, and difficulties in extracting these membrane proteins due to their short-timed activity (Fincher 2009).

Sucrose synthase catalyzes the formation of UDP-glucose and fructose from sucrose providing the immediate precursor for cellulose biosynthesis (Amor et al. 1995). Sucrose synthase can bind to the membrane rosette, a cellulose synthase complex, thereby enabling cellulose synthesis (Fujii et al. 2010). Increased sucrose synthase activity was correlated with higher cellulose content in wheat roots (Albrecht and Mustroph 2003) and poplar (Hu et al. 1999; Park et al. 2004). Overexpression of cotton sucrose synthase in hybrid poplar increased secondary cell wall cellulose content 2–6% but also cell wall crystallinity, and produced a thicker xylem of the secondary cell wall thereby improving wood quality (Coleman et al. 2009). Biomass was not affected in this transgenic poplar. However, another study demonstrated that upregulation of the enzyme produced stunted phenotypes with reduced plant height and biomass (Coleman et al. 2008). Suppression of sucrose synthase decreased cellulose content in cotton (Ruan et al. 2003). UDP-glucose pyrophosphorylase (UGPase), a second enzyme of the nucleotide-sugar interconversion pathway, independently increased cellulose in poplar but gene stacking did not provide an advantage over single sucrose synthase overexpression (Coleman et al. 2009). Simultaneous overexpression of the three sucrose pathway genes UGPase, sucrose synthase and sucrose phosphate synthase did not affect cellulose production in tobacco plants but increased plant height and delayed bud formation (Coleman et al. 2010).

Glucose is synthesized into cellulose by a rosette-like complex located at the plasma membrane harboring cellulose synthases (CESAs) and cellulose synthase like (Csl) proteins (Desprez et al. 2007; Fincher 2009; Itoh and Kimura 2001; Kimura et al. 1999; Paredez et al. 2006; Somerville 2006; Taylor et al. 2000; Taylor 2008; Yin et al. 2009). Ten CESA genes have been identified in Arabidopsis and CESA genes of rice and corn appear to be orthologous to the Arabidopsis CESAs (Penning et al. 2009; Richmond and Somerville 2000). In contrast, the CclF group of the CESA like genes has been found to be unique to grasses (Hazen et al. 2002) synthesizing a polysaccharide only found in grasses. At least three CESAs are essential for cellulose synthesis in the secondary cell wall of Arabidopsis, CESA4, -7 and -8 (Gardiner et al. 2003; Taylor et al. 2000, 2003). These are coexpressed with three CESAs (1,3,6) required for cellulose synthase in the primary cell wall

(Desprez et al. 2002, 2007; Persson et al. 2007; Robert et al. 2004; Scheible et al. 2001). Irregular xylem mutants (irx1, irx3, irx5) of Arabidopsis are defective in the catalytic subunits of the CESAs and have reduced cellulose content in the secondary cell wall (Turner and Somerville 1997; Taylor et al. 1999). The first cell wall mutant affecting conversion efficiency has recently been identified in Arabidopsis. *Ixr1-2* exhibits a mutation in a highly conserved region of CESA3 which increased efficiency of converting raw biomass into fermentable sugar to 51% of the Col-0 wild type (Harris et al. 2009). CESA3 orthologs and possible conserved domains in *Miscanthus* and other plants of interest might enable a similar mutant approach in CESAs of secondary cell walls, like CESA 4/7/8 (Harris et al. 2009).

Cellulose disassembly into simple sugars is facilitated by endoglucanases, exoglucanases and beta-glucosidases (Gray et al. 2006; Hayashi et al. 2005; Lin and Tanaka 2006). Expression of cellulose degrading enzymes *in planta* could assist in reducing saccharification costs by minimizing the need for applying external polysaccharide hydrolyzing enzymes. However, producing the appropriate amount of enzyme in the plant cell at the appropriate time will remain a challenge. These enzymes might not stay viable after an aggressive pretreatment of the plant material which is applied to degrade lignin and only limited knowledge is available on how the presence of these enzymes would affect plant growth (Abramson et al. 2010). Commonly, cellulose and hemicellulose are degraded together, which requires not one but rather a mix of different enzymes. It is also unknown how storage and activity of several enzymes in one cell compartment would influence enzyme activity. Successful expression of thermo stable bacterial cellulases in Arabidopsis, tobacco, alfalfa, potato and rice (Borkhardt et al. 2010; Dai et al. 1999; Herbers et al. 1995; Kimura et al. 2003, 2010; Ziegelhoffer et al. 1999) as well as expression of fungal cell wall degrading enzymes in potato and barley (Oomen et al. 2002; Patel et al. 2000; Sørensen et al. 2000; Yang et al. 2007) has been reported. Expressing 1,4-beta-endoglucanase E1 in corn converted stover into glucose (Ransom et al. 2007).

It needs to be considered that plants commonly activate defense mechanisms upon detecting cell wall degrading enzymes, to protect themselves against pathogen attacks. Fungi and bacteria utilize cell wall degrading enzymes in their first line of attack to be able to enter the plant cell, and plants have developed innate defense mechanisms against cell wall degradation like polygalacturonase-inhibiting proteins (Lagaert et al. 2009). If cell wall degrading enzymes are integrated into the plant cell for ease of cellulose and hemicellulose deconstruction, there might be a need to engineer the inactivation of the plant's defense mechanism after shoot senescence but not earlier during plant growth, when the plant's defense mechanism is still needed. This could be achieved by using cell type specific or inducible promoters for timely gene activation.

The application of carbohydrate binding molecules (CBM) is an additional approach to improve cell wall degradation (Abramson et al. 2010; Din et al. 1991). CBMs can bind cell wall degrading enzymes onto the surface of the polysaccharide for higher catalytic enzyme efficiency (Boraston et al. 2002; Din et al. 1991; Shoseyov et al. 2006) or are able to slide into the polymer "net" thereby increasing the surface area for better access by cellulose hydrolyzing enzymes (Abramson et al. 2010). Expansin, a similar molecule and endogenous to plants, is involved in

cell growth and flexibility and has also been considered as a cell wall "loosener" for increased saccharification (Baker et al. 2000; Cosgrove 2000, 2005; Shoseyov et al. 2006).

3.2.3 Hemicellulose

Hemicellulose synthesis is still poorly understood. Similarly to cellulose mutants, mutants with changes in enzymes of the hemicellulose pathway are severely impacted in their growth, implicating a challenge for modifying the amount of hemicelluloses as well as its properties. Hemicellulose is built of heterogeneous noncellulosic polysaccharides produced from UDP-D-glucose in the Golgi apparatus, composed of pentoses, hexoses, uronic acids, rhamnose and fucose (Scheller and Ulvskov 2010). In secondary plant cell walls, they are mostly xylans, varying between 20 and 30% dry weight in dicotylonous plants and between 40 and 50% in monocotylous plants (Vogel 2008). Arabinoglucuronoxylans and Glucuronoarabinoxylans are the main forms of xylans in grass cell walls (Ebringerová and Heinze 2000) while Xyloglucan is most abundant in dicotylous plants but is only found in minor amounts in grasses.

Hemicellulose also forms strong bonds with lignin, interfering with lignin removal and simultaneously also interfering with cellulose accessibility. Some technical pretreatments remove hemicelluloses together with lignin, followed by separate fermentation of the hemicellulose sugars into ethanol. Presently, the pentoses of the hemicelluloses cannot effectively be fermented together with the hexoses of cellulose, which is one major cost factor in the conversion of lignocellulosic biomass (Wyman 2007). A technical approach to resolve the problem is to identify or synthetically produce enzymes that can ferment both molecules in one step, defined as consolidated bioprocessing (Girio et al. 2010). On the plant side, a decrease in hemicellulose content and crystallinity combined with an increase in cellulose would be favorable for easy fermentation. If a combined fermentation becomes feasible, both cellulose and hemicellulose content should be maximized while reducing crystallinity would still be needed for improved digestion.

The enzyme UDP glucose dehydrogenase oxidizes UDP-glucose (UDP-Glc) into UDP-glucuronic acid (Tenhaken and Thulke 1996) providing the precursor for hemicelluloses as demonstrated in poplar (Johansson et al. 2002) and soybean (Tenhaken and Thulke 1996). Presumably due to their fundamental function, genes in the nucleotide-sugar conversion pathways are conserved between Arabidopsis and the grasses (Penning et al. 2009). It has been established that glycosyltransferases synthesize these molecules into hemicelluloses and studies are just starting to identify genes encoding these enzymes as well as describe their functions (Brown et al. 2005; Peña et al. 2007; Lee et al. 2010). Evidence is building that CESA-like genes involved in cellulose synthesis are also involved in hemicellulose biosynthesis (Burton et al. 2006; Cocuron et al. 2007; Dhugga et al. 2004; Doblin et al. 2009; Liepman et al. 2005; Scheller and Ulvskov 2010).

Glycosyltransferases (GTs) in poplar and Arabidopsis have been shown to be essential for proper formation of the glucuronoxylan chains and the elongation of the polymer backbone (Brown et al. 2007; Lee et al. 2010; Peña et al. 2007; Zhou et al.

2007). A gene expression study in poplar revealed 25 putative glycosyltransferases involved in secondary cell wall synthesis (Aspeborg et al. 2005). Among several gene families of Arabidopsis glycosyltransferases, the GT43, GT8 and GT47 families harbor genes essential for hemicelluloses production. The Arabidopsis genes FRAGILE FIBER 8 (FRA8), IRREGULAR XYLEM8-LIKE (IRX8), IRX9 and IRX10 have been identified to encode putative glycosyltransferases that are required for normal glucuronoxylan synthesis in secondary cell walls (Brown et al. 2009; Peña et al. 2007). The F8H glycosyltransferase is a functional paralog of FRA8 involved in glucuronoxylan biosynthesis in Arabidopsis. Although the *f8h* mutant alone did not show any detectable cell wall defects, the *f8h*/*fra8* double mutant exhibited an additional reduction in cell wall xylose level, a more severe deformation of vessels and an extreme retardation in plant growth compared with the *fra8* mutant (Wu et al. 2010). Arabidopsis IRREGULAR XYLEM9-LIKE (IRX9-L) and IRX14-LIKE (IRX14-L) encoding glycosyltransferase family 43 members have been proposed to function during xylan backbone elongation (Wu et al. 2010; Lee et al. 2010) similarly to Arabidopsis FRA8, F8H, IRX8, and PARVUS and poplar GT47C, GT8D, GT8E/8F (Aspeborg et al. 2005; Zhong et al. 2005; Brown et al. 2005, 2007; Zhou et al. 2006, 2007; Peña et al. 2007; Lee et al. 2007, 2009a, b, c).

Similarly to downregulating genes in the lignin pathway to decrease substrate production and synthesis of the polymer, a possibility would be to down regulate the above mentioned genes involved in the biosynthesis of hemicelluloses to reduce hemicelluloses or the complexity of the polymer chains and bonds. Downregulation of UDP-glucuronate decarboxylase in tobacco decreased xylan content, increased glucose content but reduced pulping yield, and left cellulose less degraded (Bindschedler et al. 2007). Lignin content was almost unchanged but lignin seemed to be more difficult to remove which might have been the cause for lower pulping yield in the mutant compared to wild type. On the other hand, downregulation of a poplar GT, PoGT47C, also decreased glucuronoxylan content but increased wood digestibility facilitated by cellulase (Lee et al. 2009b). An almost 50% xylose reduction in inflorescence stems was observed in the Arabidopsis *irx14* mutant. Xylose was drastically further reduced to about 26% in the double mutant of its homolog *irx14 i14h* which produced, however, severe growth retardation (Lee et al. 2010).

Genes for the sugar substrate conversion are highly conserved between Arabidopsis and the grasses. However, genes of the different GT families seem to have further diverged and new clades needed to be established for GTs of corn and rice in families GT47 and GT3 (Penning et al. 2009). This also suggests that GT studies for elucidating their function in grasses should rather be performed on grass model species closely related to the plant of interest.

3.2.4 Transcriptional Regulation of the Lignin, Cellulose, and Hemicellulose Pathways

Transcription factors (TFs) encode proteins with specific DNA binding domains which bind to short specific sequences located outside of coding sequences thereby activating or repressing the transcription and activity of the gene. Transcription

factors can be classified into different families according to similarities in binding domains or their regulatory functions. TFs are of particular interest for genetically modifying plants because they often control several genes within a pathway or entire pathways. Hence, the modifications of transcription factors provide the opportunity to direct entire biosynthetic pathways. An example of transcriptional regulation with significance to cell wall modification is the recently discovered set of transcription factors in Arabidopsis, poplar and Eucalyptus involved in the regulation of secondary cell wall formation. These transcription factors belong to the NAC (NAC stands for Petunia, NAM, and Arabidopsis, ATAF1, ATAF2 and CUC2) and MYB (Myeloblastosis) families and enable the activation of genes in the lignin, cellulose, and xylan pathways (Goicoechea et al. 2005; Patzlaff et al. 2003; Zhong and Ye 2007, 2010; Zhong et al. 2006, 2007a, b, 2008).

NAC TFs like VND6 (VASCULAR-RELATED NAC-DOMAIN 6) and VND7 were shown to be master regulators of xylem vessel differentiation (Kubo et al. 2005; Yamaguchi et al. 2008, 2010) and together with SND1 (SECONDARY WALL-ASSOCIATED NAC DOMAIN PROTEIN 1) and NST1 (NAC SECONDARY WALL THICKENING PROMOTING FACTOR1) modulated downstream secondary cell wall associated MYB and KNAT (KNOTTED1-LIKE HOMEODOMAIN PROTEIN) TFs, which in turn controlled the activation of secondary cell wall substrate pathways (Kubo et al. 2005; McCarthy et al. 2009; Mitsuda et al. 2007; Zhong et al. 2006, 2007a, b, 2010; Zhong and Ye 2010). EgMYB2 was cloned from Eucalyptus and colocalized with a quantitative trait locus (QTL) for lignin content (Goicoechea et al. 2005). Overexpression of EgMYB2 in tobacco changed gene expression of several monolignol pathway genes, suggesting a role for this TF in regulating lignin biosynthesis. EgMYB1 was found to be a repressor of the lignin pathway in Eucalyptus (Legay et al. 2007) and MYB46, MYB58, MYB63, and MYB85 were also found to regulate lignin biosynthesis (McCarthy et al. 2009; Zhong et al. 2007a, 2008, 2009).

The full set of TF targets regulating the lignin, cellulose, and hemicellulose pathways still needs to be elucidated. It is highly desirable to enable manipulation of carbon distribution, cell wall polymer formation and structure to optimize cellulose, hemicellulose and lignin content as well as degradability for effective processing of the plant material into bioenergy.

3.3 Modification of Plant Architecture and Stress Resistance for Biomass Yield Increase

3.3.1 Plant Biomass

It remains to be seen whether a larger economic impact on the overall economy of *Miscanthus* as a bioenergy crop will result from the optimization of cell wall composition for advanced processing, or from increasing biomass yield. Maximizing yield is the most common crop improvement goal and will surely contribute to the success

of *Miscanthus*. The widely used *M.* × *giganteus* clone yields between 10 and 30 t/ha with projections in the USA for up to 44 t/ha (Angelini et al. 2009; Clifton-Brown et al. 2001; Heaton et al. 2004a, b). *Miscanthus* utilizes the C4 photosynthetic pathway as do many of the most productive biomass plants. However, *Miscanthus* is not domesticated or locally adapted. Although natural forms produce respectable biomass yield across different geographies (Clifton-Brown et al. 2001) breeding for biomass yield promises to unlock additional potential hidden in this plant.

Biomass yield constitutes a complex combination of multiple plant traits. There are differences in yield-defining characteristics between single plants and field populations. Among the most critical traits defining plant biomass yield in *Miscanthus* are plant height, tiller number and density, and flowering time. TFs have played a major role in domestication of crops and yield increases (Century et al. 2008) and are excellent candidates for yield enhancement of undomesticated biofuel crops. Tb1 (TEOSINTE BRANCHED 1), for instance, altered branching pattern enabling the growth of high yielding corn hybrids with no branches (Doebley et al. 1997; Wang et al. 1999). Plant height reduction in dwarf rice and other grain varieties can also be attributed to TF regulation enabling much higher grain yield by reducing lodging (Peng et al. 1999; Fernandez et al. 2009). Delayed flowering in rice caused by the expression of DTH8 (DAYS TO HEADING ON CHROMOSOME 8) in a genetic background with a loss of function mutation in this gene significantly increased plant height and number of grains per panicle (Wei et al. 2010). Modulating flowering time and plant architecture to optimize tiller number and distribution, changing leaf position to increase light exposure for higher photosynthesis, increasing plant height for higher biomass yield, and increasing tolerance to biotic and abiotic stress, are possibilities for applications of TFs to increase biomass yield.

3.3.2 Plant Height

Plant height is critical for biomass yield and positive correlations between biomass yield and plant height have been observed for fodder and sweet sorghum (Ritter et al. 2008), corn (Lübberstedt et al. 1997), sugarcane (Gravois et al. 1991; Soomro et al. 2006), switchgrass (Das et al. 2004) and *Miscanthus* (Angelini et al. 2009; Jezowski 2008). Plant height is regulated by a combination of pathways related to hormone, light and floral signaling. Gibberellins and brassinosteroids are plant hormones involved in stem and cell elongation and mutants in genes of either biosynthetic pathway affect plant height. Dwarfism in rice, for instance, is caused by overexpression of genes inactivating gibberellin biosynthesis, while overexpression of genes synthesizing gibberellins lead to taller plants (Itoh et al. 2001; Li et al. 2009; Luo et al. 2006; Sasaki et al. 2003; Song et al. 2009; Tong et al. 2007; Wang et al. 2008a). A mutation in GID1 (GIBBERELIN-INSENSITIVE DWARF 1) or GID2 in rice reduced gibberellin signaling and caused dwarf phenotypes while overexpression of GID1 increased plant height (Sasaki et al. 2003; Ueguchi-Tanaka et al. 2007). A mutation in the semidwarf gene (sd1) in rice caused loss of function of the GA20 oxidase, thereby decreasing gibberellin levels and reducing plant height (Sasaki et al. 2002). In corn, several mutations have been identified causing

dwarfism, including *dwarf 8* which affected gibberellin biosynthesis (Fujioka et al. 1988). Many QTLs for plant height have been identified in sorghum and corn (Feltus et al. 2006; Pereira and Lee 1995; Quinby and Karper 1954; Lin et al. 1995). Five regions containing plant height QTLs are syntenic between sorghum and corn (Lin et al. 1995; Pereira and Lee 1995) and QTLs for flowering and plant height mapped in sorghum also corresponded to QTLs in sugarcane (Ming et al. 2002) and are good candidate locations for plant height QTLs in *Miscanthus*. QTLs for plant height, stem diameter (Atienza et al. 2003), and yield (Atienza et al. 2006) were identified from a half sib *M. sinensis* cross, however, the study could not establish a complete map based on 19 linkage groups nor could the locations (diagnosed by arbitrary-sequence RAPD primers) be aligned to other genomes.

3.3.3 Tiller Number and Branching Pattern

Angelini et al. (2009) reported a significant positive correlation between tiller number and dry matter yield in *Miscanthus*. Tiller number was also positively correlated with biomass yield in switchgrass (Das et al. 2004) but there is not always a positive correlation between tiller number and overall yield. In a separate switchgrass trial, 22 % fewer tillers in cultivar "Summer" produced 20% more biomass than "Sunburst", which was attributed to higher mass per tiller (Boe and Beck 2008).

Miscanthus is a perennial grass producing rhizomes for nutrient storage and overwintering, and tillers grow from the underground axillary buds of the rhizome. *Miscanthus* species show a range of rhizome growth behaviors from localized, clumpy rhizomes to widely spreading rhizomes. Rhizome growth considerably defines plant tillering and tiller distribution in *Miscanthus*. Studies in sorghum have identified QTLs involved in rhizome growth, rhizome number and rhizome regrowth (Paterson et al. 1995; Feltus et al. 2006). The QTLs Rhz2 and Rhz3 control rhizomatousness in rice and comparative mapping revealed that these genes are located in the same chromosomal region as a QTL affecting rhizome growth in *Sorghum propinquum* (Hu et al. 2003). The conservation of the location between these two species suggests a key regulatory role for these genes in rhizome development which seems to be maintained in grasses. Hence, a similar mapping approach might be used to identify rhizome branching related QTLs and genes in *Miscanthus*.

Tb1 affects the outgrowth of axillary buds (Doebley et al. 1995, 1997) and has played a significant role in the domestication of corn by reducing branching. In wheat, overexpression of corn Tb1 caused reduced tiller number and plant height (Lewis et al. 2008). MONOCULM (MOC1) was the first characterized gene that controls the outgrowth of rice tiller buds and the *moc1* mutant produced only one culm (Li et al. 2003). MOC1, a TF belonging to the plant-specific GRAS transcription factor family, interacted with MIP1 (MOC1 INTERACTING PROTEIN 1) and overexpression of MIP1 led to increased tillering in rice (Sun et al. 2010). It also reduced plant height which might be preferred in a grain crop (e.g., to prevent lodging) but would not be desired in a biomass crop, unless the yield reduction from decreased plant height is compensated for by higher tiller number. PROG1

(PROSTRATE GROWTH 1), a TF of the zinc finger family, has been shown to regulate branching pattern and tiller number in wild rice (Jin et al. 2008). Tiller number in rice was also found to be regulated by OsSPL14 (SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 14), another TF harboring a zinc finger binding domain (Jiao et al. 2010; Miura et al. 2010; Xie et al. 2006). Overexpression of OsSPL14 reduced tiller number but produced stronger culms and higher grain yield while suppression of OsSPL14 increased tiller number, reduced plant height and culm diameter (Jiao et al. 2010). The gene is conserved between sorghum, maize, wheat and Arabidopsis and would be a promising candidate gene for *Miscanthus* (Miura et al. 2010).

Much genotypic variation in tiller number and tiller angle has been observed in a collection of about 240 *Miscanthus* genotypes in Wales (K. Farrar, pers. comm.). Light interception is influenced by stem angle, leaf position and leaf size, which each could affect the quantity of photosynthesis and overall biomass yield. Increasing photosynthetic capacity by manipulation of plant architecture through modulation of branching patterns (Long et al. 2006; Wang et al. 2008a; Zhu et al. 2010) or leaf position, as demonstrated in rice (Li et al. 2009); are strategies that could be considered by transgenic approaches.

Tiller density is also affected by plant density in the field, and plant number per area in combination with tiller number should be optimized for highest *Miscanthus* biomass yield. Rhizome growth and tiller number vary by genotype and it is likely that different genotypes require different plant densities in the field for optimal growth and maximum biomass yield. Genes controlling tiller number and branching pattern identified in rice, corn and sorghum could be targets for optimizing tiller distribution in *Miscanthus*. This could be beneficial for optimizing tiller density in the field and increasing biomass yield, and might also reduce the invasiveness of species like *M. lutarioriparius* which develops vigorous rhizomes with a tendency to spread over large distances. While hybrids of crosses between genotypes differing in rhizome growth and vigor could produce the desired intermediate phenotype, transgenic modification of the expression of genes controlling rhizome and tiller growth could also be utilized to optimize tiller density.

3.3.4 Flowering Time

The transition to the reproductive phase terminates vegetative growth in grasses and redirects carbon assimilates toward seed production. This floral transformation is not desired in a biomass plant, but is required for breeding purposes and seed production. Flowering time variation is clearly present in *Miscanthus*, which is also a hindrance for conducting crosses within and between genotypes with differing flowering times. Synchronizing flowering of selected parents can be achieved by staggered planting, but direct regulation of flowering would allow for easier handling of crosses. Given the effect of flowering on agronomically important plant traits and seed production, regulation of flowering is a top priority for biomass plants.

Flowering is triggered by environmental signals, photoperiod, circadian clock, temperature, autonomous plant development signals and gibberellic acid response pathways (Balasubramanian et al. 2006; Dalchau et al. 2010; Greenup et al. 2009; Izawa et al. 2003; Komeda 2004; Wilczek et al. 2009) and a myriad of transcription factors are involved in flowering regulation. Mutant analysis in the long day plant Arabidopsis has enabled the discovery of genes and pathways involved in flower regulation and the CONSTANS (CO)/FLOWERING LOCUS T (FT) model has emerged as common denominator for flowering control among flowering plants (Turck et al. 2008). Further analysis in grain crops however also revealed distinctive regulatory mechanism for grasses (Colasanti and Coneva 2009; Turck et al. 2008). CO and FT are two major integrator genes of the photoperiodic and circadian clock networks controlling flowering. In Arabidopsis, GIGANTEA (GI) is a circadian clock controlled gene, which can activate the expression of CO (Fowler et al. 1999; Mizoguchi et al. 2005) which subsequently triggers FT expression while acting both as TF and coactivator (Kardailsky et al. 1999; Kobayashi et al. 1999; Tiwari et al. 2010; Yanovsky and Kay 2002). FT is synthesized in leaves and FT protein moves through the phloem to the shoot apical meristem (Abe et al. 2005; Corbesier et al. 2007) where it then interacts with the FLOWERING LOCUS D (FD) to activate APETALA1 and FRUITFUL (FUL), triggering the transformation of the apical meristem from shoot to inflorescence meristem (Abe et al. 2005; Wigge et al. 2005). SOC1 (SUPPRESSOR OF OVEREXPRESSION OF CO 1) and other genes are located downstream of FT/FD (FLOWERING LOCUS D) and are involved in flower development. Mutation in CO delays flowering by repressing the activation of FT, and activating a gene in the flowering repression pathway such as FT would render a genotype early flowering. Downregulation of SOC1 and FUL in Arabidopsis delayed or completely eliminated flowering (Melzer et al. 2008).

Genes similar to GI, CO and FT have been identified in grasses. In the short day plant rice (*Oryza sativa*), a GI ortholog has been discovered (Hayama et al. 2002) as well as a CO ortholog, Hd1 (HEADING DATE 1) (Yano et al. 2002). Hd3a in rice has been identified as an Arabidopsis FT ortholog with a similar function as in Arabidopsis (Kojiama et al. 2002). Hd3a can be activated by Hd1 (Hayama et al. 2003) and overexpression of Hd3a caused early flowering while loss of function delayed flowering. However, in contrast to Arabidopsis CO, Hd1 not only activates but could also repress Hd3a (Hayama et al. 2003). In addition, novel regulatory genes have been identified in rice which are absent in Arabidopsis, e.g., Ehd1 (EARLY HEADING DATE 1) which also regulates Hd3a expression independent of Hd1 (Doi et al. 2004; Higgins et al. 2010). DTH8 acted as a flowering suppressor under long day conditions by down-regulating Edh1 and Hd3a demonstrating possible differences in flower regulation between Arabidopsis and rice and probably other grasses (Wei et al. 2010). A CO like gene has also been identified in corn, Conz1 (MAIZE CONSTANS—LIKE) mapping to the syntenous location of Hd1 in rice, which again demonstrates the functionality of using synteny between related species to identify candidate genes (Miller et al. 2008).

TERMINAL FLOWER (TFL1) acted antagonistically to FT and repressed flower formation by maintaining an indeterminate state of the shoot apical meristem

(Bradley et al. 1996, 1997; Ratcliffe et al. 1998). Overexpression of TFL1 has been shown to extend both vegetative and reproductive phases and produced visually obvious increases in plant size (Ratcliffe et al. 1998, 1999). In Arabidopsis, mutation of TFL1 caused early flowering (Bradley et al. 1997). Genes most closely related to TFL1 belong to the CENTRORADIALIS gene family and have been identified in several plant species. CEN homologs were shown to delay flowering and vastly increased the height of tobacco plants (Amaya et al. 1999). Six CEN genes have been identified in corn (ZEA CENTRORADIALIS1–6) (Danilevskaya et al. 2008) and overexpression of ZCN2, 4 and 5 in corn caused very late flowering (Danilevskaya et al. 2010). RCN1 (RICE CENTRORADIALIS 1) and RCN2 acted similarly to TFL1 in rice and overexpression of RCN1 and RCN2 delayed flowering in Arabidopsis and rice (Nakagawa et al. 2002). In red fescue, TFL1 from perennial ryegrass was overexpressed causing delayed flowering (Jensen et al. 2001, 2004), however, culm length was reduced. This growth pattern might be desirable for obtaining highly digestible feedstock but not for maximizing biomass unless the plants are susceptible to lodging.

Temperature is an additional environmental factor affecting flowering time and FLOWERING LOCUS C (FLC) is a key gene in Arabidopsis blocking flowering by repressing FT (Searle et al. 2006). Gene expression of the flowering repressor FLC is reduced by vernalization (Michaels and Amasino 1999; Sheldon et al. 1999, 2000). VERNALIZATION genes have been identified as key genes for the vernalization in grasses (i.e.VRN1 and VRN2 in wheat, Yan et al. 2003, 2004). However, no FLC like gene has been identified in grasses, and the VRN genes of Arabidopsis are not related to the VRN genes of grasses. There are substantial differences between the vernalization pathways of Arabidopsis and the grasses while the photoperdiod and autonomous pathways are more conserved between dicots and monocots (Colasanti and Coneva 2009; Greenup et al. 2009; Izawa et al. 2003; Yan et al. 2004). Above all, the RICE INDETERMINATE 1 (RID1) gene has been proposed as a master switch for flower induction in rice. The *rid1* mutant completely prevented flowering in rice (Wu et al. 2008). No RID1 homolog was found in Arabidopsis.

Only a few additional flowering genes have been identified in corn to date. The INDETERMINATE 1 (ID1) gene is also unique to grasses with no homolog in Arabidopsis and an *id1* loss of function mutant delayed flowering in corn (Neuffer et al. 1997; Colasanti et al. 1998). It has been suggested that ID1 controls a florigenic signal and most likely acts in the autonomous pathway (Colasanti et al. 1998; Colasanti and Sundaresan 2000; Kozaki et al. 2004). DLF1 (DELAYED FLOWERING 1) is another flowering gene identified in corn acting downstream of ID1 with loss of function also delaying flowering (Neuffer et al. 1997; Muszynski et al. 2006). Both ID1 and DFL1 are transcription factors (Kozaki et al. 2004; Muszynski et al. 2006). LEAFY1 has also been found to delay flowering in corn but is distinct from the AtLeafy1 (Neuffer et al. 1997; Passas and Poethig 1993).

The identification of flowering time genes in corn and sugarcane is still lagging behind other plants like Arabidopsis and rice. In addition to mutant analysis, QTL analysis has been the source for identifying genes in Arabidopsis and similarly, QTL analysis for flowering time for both corn and sugarcane has been and will further be

employed to identify genes related to flowering time. Studies in corn have revealed several QTLs for flowering (Chardon et al. 2004; Buckler et al. 2009) but only a few of these genes have been cloned. In addition to the above described DLF1, ID1 and DWARF8, the VGT1 (VEGETATIVE TO GENERATIVE TRANSITION 1) QTL has been cloned and ZmRap2.7 (APETALA2-like) has been identified as a flowering repressor located downstream of VGT1 (Salvi et al. 2007). Overexpression of ZmRap2.7 delayed flowering in corn and also increased plant height (Salvi et al. 2007). Four main QTL regions controlling photoperiodic response were identified on corn chromosomes 1, 8, 9 and 10 along with several candidate genes for flowering time (Coles et al. 2010). These four QTLs have also been identified in a nested association mapping study in corn which concluded that differences in flowering time in corn are the result of many QTLs with small effects (Buckler et al. 2009). This is in contrast to few QTLs having large effects on flowering time in Arabidopsis and rice (Koornneef et al. 1998; Yano and Sasaki 1997). However, in a study of flowering time of Arabidopsis in natural populations, additive effects of many small QTLs were also found (Brachi et al. 2010). Six flowering time-related genes, Ma1–Ma6 (MATURITY), responsible for modulating flowering time have been identified in sorghum (Quinby and Karper 1945; Quinby 1966, 1974) and Ma1 and Ma3 have been mapped (Childs et al. 1997; Klein et al. 2008). A crossing strategy in combination with molecular markers for Ma5 and Ma6 has enabled the selection of late and early flowering sorghum genotypes (Mullet et al. 2010). Ma5 and Ma6 in the dominant state caused late flowering, whereas the recessive alleles permitted sorghum to flower early (Rooney and Aydin 1999). Similarly to Arabidopsis and rice, QTLs for flowering time in sorghum also show large effects (Lin et al. 1995). While Arabidopsis, rice and sorghum are self fertile, corn is more frequently an outcrosser, suggesting that differences in the effect of QTLs on flowering time could result from the differences in mating systems (Buckler et al. 2009). *Miscanthus* is an outcrossing species with variation in flowering time related to photoperiod, i.e., *M. sacchariflorus* collected in tropical latitudes needs short days for flower induction while *M. sacchariflorus* robustus seems to be adapted to flowering under long day conditions. A positive correlation between flowering time and plant height has been demonstrated in several plants. Hence, it is not surprising that some QTLs for flowering time are colocalizing with QTLs for plant height as has been established in corn and sorghum (Coles et al. 2010; Lin et al. 1995; Murray et al. 2008; Ritter et al. 2008).

Manipulation of flowering time in *Miscanthus* is highly desirable. Photoperiod sensitive and insensitive genotypes could be selected and grown at the appropriate latitude to achieve late flowering and high biomass yield. However, direct flowering regulation would permit the few most elite genotypes to be utilized in many potential growing regions, regardless of photoperiod. Differences between monocots and dicots in flowering pathways and gene functions emphasize the importance of focusing on grass species as model plants for understanding flowering in *Miscanthus*. Synteny between rice, sorghum and sugarcane is expected to extend to *Miscanthus*, and could expedite identification of flowering-related QTLs and

genes in *Miscanthus*. This knowledge can then be applied for adjusting flowering time to the *Miscanthus* product goal (ornamental use, biomass, or seed production), and latitude.

3.3.5 Biotic Stress

Miscanthus is closely related to sugarcane, sorghum and corn and therefore these crops might be expected to share common pests and diseases. *Miscanthus* genotypes are highly heterozygous polyploids with built-in diversity which could prove sufficient to provide many disease resistance genes as they become necessary. *Miscanthus* has been used as a source of disease resistance genes which have been introgressed into sugarcane (Chen and Lo 1989; Miller et al. 2005). *Miscanthus* has been grown as an ornamental plant in a wide range of geographies and in larger plantations as biomass resource in the UK, Germany and Switzerland. Serious pest and pathogen outbreaks have not yet been observed, but pathogens and diseases have been reported in *Miscanthus* in its native range, on ornamentals, and in field trials. In addition to the above mentioned observation of *Fusarium* and the Asiatic pink stem borer on *Miscanthus*, *F. culmorum* and *F. graminearum* have been shown to reduce survival and both above- and below-ground biomass accumulation of young *Miscanthus* plants (Gossmann 2000; Thinggaard 1997). These two *Fusarium* species are important pathogens of wheat and barley causing seedling blight, foot rot, and *Fusarium* head blight, and epidemic outbreaks have occurred in North America (Windels 2000). *Miscanthus* blight has been observed on ornamental *M. sinensis* (O'Neill and Farr 1996). *Miscanthus* streak and barley yellow dwarf virus (BYDV) were found on wild *M. sacchariflorus* in Japan (Yamashita et al. 1985) and on *M. sacchariflorus* in the UK (Christian et al. 1994), respectively. BYDV is transmitted by *Rhopalosiphum maidis* (corn leaf aphid) and reduced above ground biomass of *Miscanthus* by 23% (Huggett et al. 1999). The corn leaf aphid is also a vector for corn dwarf mosaic virus and mosaic virus of sugarcane, causing damage in both crops as well as in sorghum and other grasses (Teakle et al. 1989). Sugarcane mosaic virus has recently been reported on *Miscanthus* (Agindotan et al. 2010). Mealybugs have also been observed on *Miscanthus* (Halbert and Remaudiere 2000) as has the Northern root knot nematode, *Meloidogyne hapla*, which is the most important pathogen of flowering annual ornamentals (LaMondia 1996).

Except for genetically engineered virus resistance in potato, plum, squash and papaya, disease resistance in commercial crops has been mostly achieved by using naturally occurring variation. Single major gene resistance is effective in combating a particular pathogen but is sometimes overcome by the pathogen, i.e., in *B. napus*, where resistance to *Leptosphaeria maculans* was overcome within 3 years (Sprague et al. 2006). While there are several options to provide broad spectrum or non host disease resistance, the transgenic alteration of expression of TFs involved in the regulation of one or more defense pathways may provide opportunities for highly desired broad-spectrum disease resistance. Immune defense in plants is triggered by either effector proteins of an attacking plant pathogen, which are recognized by

plant resistance genes (ETI) or by certain molecular signatures of pathogens (PAMPS, PTI). Plant immune defense involves the activation of salicylic (SA), jasmonic (JA) and ethylene (ET) pathways. Biotrophic and necrotrophic pathogens commonly elicit the SA and JA/ET pathways, respectively, whereby SA is usually regulated antagonistically to the JA/ET pathway (Kunkel and Brooks 2002). In recent years, the role of WRKY TFs in the regulation of plant defense responses has become more evident (Pandey and Somssich 2009) and AtWRKY70 has been identified as an integrator gene at the junction of the SA and the JA/ET pathway (Li et al. 2004). AtWRKY70 has been shown to be required for full RPP4-mediated disease resistance and basal defense against *Hyaloperonospora parasitica* in Arabidopsis (Knoth et al. 2007). More than 100 WRKYs have been identified in the rice genome (Ross et al. 2007) and at least eight WRKYs were involved in defense responses to pathogens (Jang et al. 2010). Overexpression of WRKY13 in rice enhanced resistance to *Xanthomonas oryza* and *Magnaportha grisea* (Qiu et al. 2007). WRKY89 overexpression enhanced resistance to *M. grisea* and white-backed leaf hopper (Wang et al. 2007), regulating resistance against several different pests. In barley, WRKY1 and WRKY2 were associated with the avirulence-R-gene triggered defense response against powdery mildew (Shen et al. 2007). A total of 68 WRKYs have been identified in the sorghum genome (Pandey and Somssich 2009) but their function has not yet been elucidated.

ERF (ethylene response factor) genes, which belong to the AP2 class of transcription factors that are unique to plants, have also been implicated as important regulators of plant defense responses (Gutterson and Reuber 2004). *Pti4* of tomato confers resistance to *Erysiphe orontii* and increased tolerance to the bacterial pathogen *Pseudomonas syringae* pv *tomato* in tomato (Gu et al. 2002). Homologous ERFs have also been successfully tested in tobacco (Park et al. 2001; Zuo et al. 2007) and wheat (Xu et al. 2007). Similarly, ERF1 can provide enhanced resistance to several pathogens, including the necrotrophic fungi *Botrytis cinerea* and *Plectosphaerella cucumerina*, and the soilborne fungi *Fusarium oxysporum* sp. *conglutinans* and *F. oxysporum* f. sp. *lycopersici* (Berrocal-Lobo et al. 2002). The ability of TFs to provide broad spectrum disease resistance enhances the possibility to become an economically viable option for genetically engineered disease resistance.

3.3.6 Abiotic Stress

Miscanthus will be grown as a non food crop and it is expected that marginal land will be used for it, leaving fertile soils in productive climates for food production. Hence, *Miscanthus* with increased tolerance to suboptimal growing conditions like drought, cold and low soil fertility will be required, making these traits long-term breeding goals.

Water use efficiency of *M.* × *giganteus* tested in a temperate climate is comparable to other C4 plants but considerably higher than most C3 plants (Beale et al 1999). Nevertheless, biomass yield is positively correlated with water supply during

the growing season (Heaton et al 2004a). C4 crops are among the highest yielding crops but require higher temperatures (above 15°C) for optimal photosynthesis than C3 plants. However, *M.* × *giganteus* has been shown to maintain photosynthesis at relatively low temperatures, due to higher chloroplastic pyruvate phosphate dikinase (PPDK) RNA accumulation in response to cold (Wang et al. 2008b). The native habitats of *Miscanthus* species varies over many degrees of latitude and it is presently not known if this ability to maintain high photosynthetic efficiencies at low temperatures is widely distributed among *Miscanthus* species and genotypes, or unique to only a few that happen to be well studied. *Miscanthus* winter kill has been observed in Europe (Clifton-Brown and Lewandowski 2000; Pude et al. 1996) which caused significant economic loss.

The effect of nitrogen on *Miscanthus* yield is inconsistent. While most studies over a wide range of geographies have shown no significant growth increase with added nitrogen (Heaton et al. 2004a; Christian et al. 2008), one study does report *Miscanthus* yield increase with increasing amounts of added nitrogen under irrigation (Ercoli et al. 1999). During senescene, *Miscanthus* translocates nutrients from above ground biomass into the rhizomes. These nutrients are retranslocated into the aerial portions of the plant during the next growing season. This could be among the reasons that no effect of nitrogen fertilization has been observed in most studies. Timely senescence is not only desired to ensure nutrient relocation but for quality of the harvest material. High nutrient content in the leaves and stems can complicate further processing, i.e., causing high ash content.

In addition to their newly discovered functions in biotic stress responses, WRKY TFs have also been involved in abiotic stress responses. Overexpression of rice WRKY45 in Arabidopsis increased salt and drought tolerance, and disease resistance (Qiu and Yu 2009), emphasizing their role in activating multiple stress signaling pathways both for biotic and abiotic stress. WRKYs are also involved in senescence (Ay et al. 2009; Jing et al. 2009; Uelker et al. 2007) and overexpression of rice WRKY23 accelerated senescence in Arabidopsis (Jing et al. 2009). Regulation of senescence could be an effective trait in *Miscanthus* for regions where senescence is delayed and *Miscanthus* cannot dry in time for the harvest. Senescence before harvest is desired as valuable nutrients are preserved as is described above. NAC TFs have also been shown to regulate abiotic stress responses. Overexpression of OsNAC6 in rice resulted in drought and salt stress tolerance, but yield was negatively affected while no effect on yield was observed in SNAC1 and SNAC2 transgenic rice expressing enhanced drought and salt tolerance, and cold and salt tolerance, respectively (Hu et al. 2006, 2008).

TFs of the CBF/DREB (C-repeat/dehydration responsive element binding factor) family have long been proven to regulate activation of plant responses to abiotic stresses like cold, salinity and drought. Cold tolerance in *Miscanthus* would be advantageous for facilitating early spring planting as well as for enabling nutrient translocation into the rhizome even after freezing periods. Homologs of Arabidopsis DREB genes have been identified in several grasses, activating similar stress responses (Wang et al. 2008c; Qin et al. 2004). Overexpression of DREB1A from wild barley in bahiagrass under a stress inducible promoter provided enhanced

resistance to severe salt and drought stress (James et al. 2008). The rice ortholog OsDREB1A overexpressed in rice conferred drought, salt and low temperature tolerance (Ito et al. 2006; Shinozaki and Yamaguchi-Shinozaki 2007). Overexpression of AtDREB1A in tobacco provided drought and cold tolerance and the stress inducible promoter *rd29A* was used to minimize the negative effect on plant growth (Kasuga et al. 2004). Most recently it has also been shown that CBF2 in Arabidopsis not only induced cold tolerance but also delayed leaf senescence (Sharabi-Schwager et al. 2010). DREB2A was cloned from pearl millet responding to cold stress and provided stress resistance in tobacco (Agarwal et al. 2007, 2010).

4 Constraints

4.1 Transformation

There is still a need to research the potential to genetically engineer *Miscanthus*. Although the protocol for transformation described above is reliable, it has not been applied to all *Miscanthus* genotypes, and there is the possibility that genotypes might exist, or might emerge from a breeding program, which do not transform well, or at all. Which promoters to use to drive strong or weak constitutive, tissue-specific, or developmentally specific expression must still be determined. There is also much that needs to be learned about how transgenes might perform in *Miscanthus*. Are they normally transmitted to progeny? Will silencing be a problem? The appearance of obvious somaclonal variants has been rare in the *Miscanthus* plants regenerated from ecallus; however, no large scale trials have been done to investigate the frequencies of variants which might be more subtle such as those reducing yield or resistance to pests or stress. Finally, transgenic technology will find its best use if integrated into a breeding program, with very careful selection of the variety or breeding line to be transformed. At this early stage, such data is only beginning to be collected.

Transgenes have the potential to improve the characteristics of the plant into which they are inserted, and therefore transgenic approaches are often integrated into breeding programs aimed at multiple aspects of crop improvement. For crops in which there are well known elite cultivars, the choice of genotype to receive the transgenes is often clear, especially if the elite cultivar lacks a trait that can be improved by the presence of a transgene. For crops which have received little or no prior breeding attention, a great deal of improvement may be possible by the application of classical breeding techniques, which could be assisted by molecular marker technology. This is the case with *Miscanthus*, a crop which has received very little attention from plant breeders to date, and the little breeding that has been done has focused on ornamental rather than biomass traits.

Once a transformation protocol has been developed, the production of new cultivars by insertion of transgenes will still require many time consuming steps: (1) production of many insertion events by application of the transformation protocol; (2) selection of one or a few insertion events of choice, including gene expression and

regulatory considerations, (3) multiple years of field trials in which the characteristics of the proposed new variety are evaluated, including but not limited to the performance of the transgenic trait; (4) de-regulation of one or more insertion events, eventually resulting in a transgenic variety or varieties which can be grown without the need for permits or other regulation; (5) bulking and further trialing of the plants containing the event, and finally release of a new cultivar. Each of these steps will require multiple years of work. In that active breeding programs exist and early progress should be rapid, transgenic forms of todays *Miscanthus* varieties (e.g., *M.* × *giganteus* var Illinois) are likely to be hopelessly out of date by the time that they could be commercialized. Improvement of *Miscanthus* via transformation technology should focus on the improvement of advanced breeding lines by insertion of transgenes coding for high value traits.

4.2 Invasiveness

Invasiveness and gene transfer from genetically modified crops into any population, crop or weed, are major concerns and need to be addressed before large scale production of transgenic *Miscanthus* takes place. *Miscanthus* is a perennial, outcrossing plant native to Asia and has no native relatives in the USA. However, outcrossing with ornamental *M. sinensis* planted throughout the USA is a possibility. *Miscanthus* can propagate via rhizomes and seed but some species are sterile. Sorghum and sugarcane are most closely related to *Miscanthus* and crosses can be made between *Miscanthus* and the two species. However it is not clear if natural hybridization can occur.

The varieties of *Miscanthus* that probably will be commercialized first are triploid and sterile. These varieties must be vegetatively propagated (usually from rhizomes), and they are therefore slow and expensive to establish. Nonseeded varieties of *Miscanthus* are not considered potential weed risks (Barney and DiTomaso 2008). It is anticipated that seeded *Miscanthus* varieties will ultimately replace the sterile varieties due to the need to establish very large acreage, and the impracticality of vegetatively propagating the number of plants that would be necessary. Seeded varieties will be much more practical as a means to establish large acreages of *Miscanthus*.

A concern for seeded varieties of *Miscanthus* is the possibility that they could become problem weeds. This concern would be heightened by the presence of transgenes; especially transgenes intended to increase the fitness of a transgenic plant or which might compromise the ability to eradicate it (i.e., herbicide resistance). Many of the transgenes under consideration (e.g., stress resistance, yield, nitrogen use efficiency, etc.) could be seen as increasing the fitness of the transgenic variety, and therefore may increase the likelihood of creating problem weeds.

Miscanthus will be grown largely for biomass, which from a standpoint of transgene containment is advantageous over grain crops where flowering and seed production is the ultimate product goal. In contrast, flowering is unwanted in *Miscanthus*, because it ends culm growth and thus biomass production. Hence, the ideal *Miscanthus* plant does not flower, rendering gene flow impossible. As described

above, modification of flowering time for increased biomass yield is an important breeding target in *Miscanthus* and should enable the design of late or non flowering genotypes which would significantly reduce the probability of gene flow. The presently cultivated *M.* × *giganteus* is thought to be sterile and present no threat of gene transfer by seeds if genetically modified. Several strategies are available to prevent the release of an invasive seeded *Miscanthus* genotype. These strategies include the use of the self incompatibility characteristic naturally present within most *Miscanthus* species, selection of genotypes with flowering times so late that they are unlikely to produce seeds prior to the end of the growing season, and transgenic strategies. Several molecular strategies have been developed to prevent pollen and seed dispersal. Male sterility, cytoplasmic male sterility (CMS), seed sterility, plastid transformation, maternal inheritance and others (Daniell 2002; Kausch et al. 2010; Lee and Natesan 2006) could be engineered in *Miscanthus* to ensure that no transgene can escape into other plants.

Misanthus produces underground rhizomes, and rhizome growth behavior varies significantly between species. While the rhizomes in some species develop only around the originally planted rhizome or plantlet (i.e., in *M.* × *giganteus*), rhizomes of other species like *M. lutarioriparius* extend rather far. In addition, *M. sacchariflorus* and *M. lutarioriparius* can develop shoots and roots from axillary buds of stem cuttings similar to sugarcane. These plant characteristics are a feature of several invasive species and can be addressed by breeding for an intermediate rhizome phenotype or plant tillers without axillary buds.

The introduction of genetically engineered herbicide resistance in crops has greatly changed weed management in crops and the area of herbicide application has been dramatically increased (Dill et al. 2008). Besides the obstacle of eradicating herbicide resistant *Miscanthus*, repeated application of only one herbicide has expectedly caused the development of resistant weed species. For glysophate, 19 resistant weeds have been reported so far (www.weedscience.org). To counteract this development it will be necessary to engineer multiple-herbicide crop resistance (Green et al. 2008).

4.3 Regulatory

Miscanthus is a strong contender for genetic modification by biotechnology. Regulatory processing is necessary as for any other modified crop. However, given that *Miscanthus* is not used as a feed or food crop, we do not consider it necessary to use the laborious and time-consuming procedures for production of transgenic varieties lacking selectable markers. In addition, regulatory processing could likely be faster, saving time and money and accordingly lowering the economic value threshold necessary to justify introduction of genetically modified *Miscanthus* as a competitive bioenergy plant. For regulatory approval in the USA it is of benefit that *Miscanthus* is not native, avoiding the outcross of transgenes to natural populations which remains a risk for switchgrass.

References

Abe M, Kobayashi Y, Yamamoto S, Daimon Y, Yamaguchi A, Ikeda Y, Ichinoki H, Notaguchi M, Goto K, Araki T (2005) FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. Science 309:1052–1056

Abramson M, Shoseyov O, Shani Z (2010) Plant cell wall reconstruction toward improved lignocellulosic production and processability. Plant Sci 178:61–72

Agarwal P, Agarwal PK, Nair S, Sopory SK, Reddy MK (2007) Stress-inducible DREB2A transcription factor from *Pennisetum glaucum* is a phosphoprotein and its phosphorylation negatively regulates its DNA-binding activity. Mol Genet Genomics 277:189–198

Agarwal P, Agarwal PK, Joshi AJ, Sopory SK, Reddy MK (2010) Overexpression of PgDREB2A transcription factor enhances abiotic stress tolerance and activates downstream stress-responsive genes. Mol Biol Rep 37:1125–1135

Agindotan BO, Ahonsi MO, Domier LL, Gray ME, Bradley CA (2010) Application of sequence-independent amplification (SIA) for the identification of RNA viruses in bioenergy crops. J Virol Methods 169:119–128. doi:10.1016/j.jviromet.2010.07.008

Albrecht G, Mustroph A (2003) Localization of sucrose synthase in wheat roots: increased in situ activity of sucrose synthase correlates with cell wall thickening by cellulose deposition under hypoxia. Planta 217:252–260

Ali MB, Singh N, Shohael AM, Hahn EJ, Paek K-Y (2006) Phenolics metabolism and lignin synthesis in root suspension cultures of *Panax ginseng* in response to copper stress. Plant Sci 171:147–154

Amaya I, Ratcliffe OJ, Bradley DJ (1999) Expression of CENTRORADIALIS (CEN) and CEN-like genes in tobacco reveals a conserved mechanism controlling phase change in diverse species. Plant Cell 11:1405–1417

Amor Y, Haigler C, Johnson S, Wainscott M, Delmer DP (1995) A membrane-associated form of sucrose synthase and its potential role in synthesis of cellulose and callose in plants. Proc Natl Acad Sci USA 92:9353–9357

Angelini LG, Ceccarini L, Nassi o Di Nasso N, Bonari E (2009) Comparison of *Arundo donax* L. and *Miscanthus* × *giganteus* in a long-term field experiment in Central Italy: analysis of productive characteristics and energy balance. Biomass Bioener 33:635–643

Archer T, Patrick C, Schuster G, Cronholm G, Bynum ED Jr, Morrison WP (2001) Ear and shank damage by corn borers and corn earworm to four events of *Bacillus thuringiensis* transgenic maize. Crop Prot 20:139–144

Aspeborg H, Schrader J, Coutinho PM, Stam M, Kallas A, Djerbi S, Nilsson P, Denman S, Amini B, Sterky F, Master E, Sandberg G, Mellerowicz E, Sundberg B, Henrissat B, Teeri TT (2005) Carbohydrate-active enzymes involved in the secondary cell wall biogenesis in hybrid aspen. Plant Physiol 137:983–997

Aswath CR, Mo SY, Kim DH, Park SW (2006) Agrobacterium and biolistic transformation of onion using non-antibiotic selection marker phosphomannose isomerase. Plant Cell Rep 25:92–99

Atienza SG, Satovic Z, Petersen KK, Dolstra O, Martín A (2003) Identification of QTLs influencing agronomic traits in *Miscanthus sinensis* Anderss. I. Total height, flag-leaf height and stem diameter. Theor Appl Genet 107:123–129

Atienza SG, Satovic Z, Petersen KK, Dolstra O, Martín A (2006) Identification of QTLs associated with yield and its components in *Miscanthus sinensis* Anderss. Euphytica 132:353–361

Ay N, Irmler K, Fischer A, Uhlemann R, Reuter G, Humbeck K (2009) Epigenetic programming via histone methylation at WRKY53 controls leaf senescence in *Arabidopsis thaliana*. Plant J 58:333–346

Baker JO, King MR, Adney WS, Decker SR, Vinzant TB, Lantz SE, Nieves RE, Thomas SR, Li LC, Cosgrove DJ, Himmel ME (2000) Investigation of the cell-wall loosening protein expansin as a possible additive in the enzymatic saccharification of lignocellulosic biomass. Appl Biochem Biotechnol 84–86:217–223

Balasubramanian S, Sureshkumar S, Lempe J, Weigel D (2006) Potent induction of *Arabidopsis thaliana* flowering by elevated growth temperature. PLoS Genet 2:e106

Barney JN, DiTomaso JM (2008) Nonnative species and bioenergy: are we cultivating the next invader? Bioscience 58:64–70

Barrière Y, Méchin V, Lafarguette F, Manicacci D, Guillon F, Wang H, Lauressergues D, Pichon M, Bosio M, Tatout C (2009) Toward the discovery of maize cell wall genes involved in silage quality and capacity to biofuel production. Maydica 54:161–198

Beale CV, Morison JIL, Long SP (1999) Water use efficiency of C4 perennial grasses in a temperate climate. Agric For Meteorol 96:103–115

Berrocal-Lobo M, Molina A, Solano R (2002) Constitutive expression of ETHYLENE-RESPONSE-FACTOR1 in Arabidopsis confers resistance to several necrotrophic fungi. Plant J 29:23–32

Bhuiyan NH, Selvaraj G, Wei Y, King J (2009) Role of lignification in plant defense. Plant Signal Behav 4:158–159

Bindschedler LV, Tuerck J, Maunders M, Ruel K, Petit-Conil M, Danoun S, Boudet AM, Joseleau JP, Bolwell GP (2007) Modification of hemicellulose content by antisense down-regulation of UDP-glucuronate decarboxylase in tobacco and its consequences for cellulose extractability. Phytochemistry 68:2635–2648

Boe A, Beck DL (2008) Yield components of biomass in switchgrass. Crop Sci 48:1306–1311

Boraston AB, Ghaffari M, Warren RAJ, Kilburn DG (2002) Identification and glucan-binding properties of a new carbohydrate-binding module family. Biochem J 361:35–40

Borkhardt B, Harholt J, Ulvskov P, Ahring BK, Jørgensen B, Brinch-Pedersen H (2010) Autohydrolysis of plant xylans by apoplastic expression of thermophilic bacterial endo-xylanases. Plant Biotechnol 8:363–374

Brachi B, Faure N, Horton M, Flahauw E, Vazquez A, Nordborg M, Bergelson J, Cuguen J, Roux F (2010) Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. PLoS Genet 6:e1000940

Bradley D, Carpenter R, Copsey L, Vincent C, Rothstein S, Coen E (1996) Control of inflorescence architecture in *Antirrhinum*. Nature 379:791–797

Bradley D, Ratcliffe O, Vincent C, Carpenter R, Coen E (1997) Inflorescence commitment and architecture in Arabidopsis. Science 275:80–83

Brown DM, Zeef LA, Ellis J, Goodacre R, Turner SR (2005) Identification of novel genes in Arabidopsis involved in secondary cell wall formation using expression profiling and reverse genetics. Plant Cell 17:2281–2295

Brown DM, Goubet F, Wong VW, Goodacre R, Stephens E, Dupree P, Turner SR (2007) Comparison of five xylan synthesis mutants reveals new insight into the mechanisms of xylan synthesis. Plant J 52:1154–1168

Brown PJ, Rooney WL, Franks C, Kresovich S (2008) Efficient mapping of plant height quantitative trait loci in a sorghum association population with introgressed dwarfing genes. Genetics 180:629–637

Brown DM, Zhang Z, Stephens E, Dupree P, Turner SR (2009) Characterization of IRX10 and IRX10-like reveals an essential role in glucuronoxylan biosynthesis in Arabidopsis. Plant J 57:732–746

Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC, Goodman MM, Harjes C, Guill K, Kroon DE, Larsson S, Lepak NK, Li H, Mitchell SE, Pressoir G, Peiffer JA, Rosas MO, Rocheford TR, Romay MC, Romero S, Salvo S, Sanchez Villeda H, da Silva HS, Sun Q, Tian F, Upadyayula N, Ware D, Yates H, Yu J, Zhang Z, Kresovich S, McMullen MD (2009) The genetic architecture of maize flowering time. Science 325:714–718

Burton RA, Wilson SM, Hrmova M, Harvey AJ, Shirley NJ, Medhurst A, Stone BA, Newbigin EJ, Bacic A, Fincher GB (2006) Cellulose synthase-like CslF genes mediate the synthesis of cell wall (1,3;1,4)-β-D-glucans. Science 311:1940–1942

Casas AM, Kononowicz K, Zehr UB, Tomes DT, Axtell JD, Butler LG, Bressan RA, Hasegawa PM (1993) Transgenic sorghum plants via microprojectile bombardment. Proc Natl Acad Sci USA 90:11212–11216

Castro BA, Leonard BR, Riley TJ (2004) Management of feeding damage and survival of southwestern corn borer and sugarcane borer (Lepidoptera: Crambidae) with *Bacillus thuringiensis* transgenic field corn. J Econ Entomol 97:2106–2116

Catangui MA, Berg RK (2006) Western bean cutworm, *Striacosta albicosta* (Smith) (Lepidopteran: Noctuidae), as a potential pest of transgenic Cry1Ab *Bacillus thuringiensis* corn hybrids in South Dakota. Environ Entomol 35:1439–1452

Century K, Reuber TL, Ratcliffe OJ (2008) Regulating the regulators: the future prospects for transcription-factor-based agricultural biotechnology products. Plant Physiol 147:20–29

CERA (2012) GM crop database. Center for Environmental Risk Assessment (CERA), ILSI Research Foundation, Washington D.C. http://cera-mc.org/index.php?action=gm_crop_database

Chardon F, Virlon B, Moreau L, Falque M, Joets J, Decousset L, Murigneux A, Charcosset A (2004) Genetic architecture of flowering time in maize as inferred from quantitative trait loci meta-analysis and synteny conservation with the rice genome. Genetics 168:2169–2185

Chen F, Dixon RA (2007) Lignin modification improves fermentable sugar yields for biofuel production. Nat Biotechnol 25:759–761

Chen YH, Lo CC (1989) Disease resistance and sugar content in *Saccharum–Miscanthus* hybrids. Taiwan Sugar 36:912

Chen T-S, Lam L, Chen S-C (1985) Somatic embryogenesis and plant regeneration from cultured young inflorescences of *Oryza sativa* L. (rice). Plant Cell Tissue Organ Culture 4:51–54

Cherney JH, Axtell JD, Hassen MM, Anliker KS (1998) Forage quality characterization of a chemically-induced brown-midrib mutant in pearl-millet. Crop Sci 28:783–787

Childs KL, Miller FR, Cordonnier-Pratt MM, Pratt LH, Morgan PW, Mullet JE (1997) The sorghum photoperiod sensitivity gene, Ma3, encodes a phytochrome B. Plant Physiol 113:611–619

Christian DG, Lamptey JNL, Forde SMD, Plumb RT (1994) First report of barley yellow dwarf luteovirus on *Miscanthus* in the United Kingdom. Eur J Plant Pathol 100:167–170

Christian DG, Riche AB, Yates NE (2008) Growth, yield and mineral content of *Miscanthus* × *giganteus* grown as a biofuel for 14 successive harvests. Ind Crop Prod 28:320–327

Clifton-Brown JC, Lewandowski I (2000) Overwintering problems of newly established *Miscanthus* plantations can be overcome by identifying genotypes with improved rhizome cold tolerance. New Phytol 148:287–294

Clifton-Brown JC, Lewandowski I, Andersson B, Basch G, Christian DG, Bonderup-Kjeldsen J, Jørgensen U, Mortensen V, Riche AB, Schwarz KU, Tayebi K, Teixeira F (2001) Performance of 15 *Miscanthus* genotypes at five sites in Europe. Agron J 93:1013–1019

Cocuron JC, Lerouxel O, Drakakaki G, Alonso AP, Liepman AH, Keegstra K, Raikhel N, Wilkerson CG (2007) A gene from the cellulose synthase-like C family encodes a beta-1,4 glucan synthase. Proc Natl Acad Sci USA 104:8550–8555

Colasanti J, Coneva V (2009) Mechanisms of floral induction in grasses: something borrowed, something new. Plant Physiol 149:56–62

Colasanti J, Sundaresan V (2000) 'Florigen' enters the molecular age: long distance signals that cause plants to flower. Trends Biochem Sci 25:236–240

Colasanti J, Yuan Z, Sundaresan V (1998) The indeterminate gene encodes a zinc finger protein and regulates a leaf-generated signal required for the transition to flowering in maize. Cell 93:593–603

Coleman HD, Samuels AL, Guy R, Mansfield SD (2008) Perturbed lignification impacts tree growth in hybrid poplar—a function of sink strength, vascular integrity and photosynthetic assimilation. Plant Physiol 148:1229–1237

Coleman HD, Yan J, Mansfield SD (2009) Sucrose synthase affects carbon partitioning to increase cellulose production and altered cell wall ultrastructure. Proc Natl Acad Sci USA 106: 13118–13123

Coleman HD, Beamish L, Reid A, Park JY, Mansfield SD (2010) Altered sucrose metabolism impacts plant biomass production and flower development. Transgenic Res 19:269–283

Coles ND, McMullen MD, Balint-Kurti PJ, Pratt RC, Holland JB (2010) Genetic control of photoperiod sensitivity in maize revealed by joint multiple population analysis. Genetics 184:799–812

Corbesier L, Vincent C, Jang S, Fornara F, Fan Q, Searle I, Giakountis A, Farrona S, Gissot L, Turnbull C, Coupland G (2007) FT protein movement contributes to long-distance signaling in floral induction of *Arabidopsis*. Science 316:1030–1033

Cosgrove DJ (2000) Loosening of plant cell walls by expansions. Nature 407:321–326
Cosgrove DJ (2005) Growth of the plant cell wall. Nat Rev Mol Cell Biol 6:850–861
Dai Z, Hooker B, Quesenberry R, Gao J (1999) Expression of *Trichoderma reesei* exocellobiohydrolase I in transgenic tobacco leaves and calli. Appl Biochem Biotechnol 79:689–699
Dalchau N, Hubbard KE, Robertson FC, Hotta CT, Briggs HM, Stan GB, Goncalves JM, Webb AAR (2010) Correct biological timing in Arabidopsis requires multiple light-signaling pathways. Proc Natl Acad Sci USA 107:13171–13176
Daniell H (2002) Molecular strategies for gene containment in transgenic crops. Nature Biotechnology 20:581–586
Danilevskaya ON, Meng X, Hou Z, Ananiev EV, Simmons CR (2008) A genomic and expression compendium of the expanded PEBP gene family from maize. Plant Physiol 146:250–264
Danilevskaya ON, Meng X, Ananiev EV (2010) Concerted modification of flowering time and inflorescence architecture by ectopic expression of TFL1-like genes in maize. Plant Physiol 153: 238–251
Das MK, Fuentes RG, Taliaferro CM (2004) Genetic variability and trait relationships in Switchgrass. Crop Sci 44:443–448
De Cesare M, Hodkinson TR, Barth S (2010) Chloroplast DNA markers (cpSSRs, SNPs) for *Miscanthus*, *Saccharum* and related grasses (Panicoideae, Poaceae). Mol Breed. doi:10.1007/s11032-010-9451-z
Desprez T, Vernhettes S, Fagard M, Refrégier G, Desnos T, Aletti E, Py N, Pelletier S, Höfte H (2002) Resistance against herbicide isoxaben and cellulose deficiency caused by distinct mutations in same cellulose synthase isoform CESA6. Plant Physiol 128:482–490
Desprez T, Juraniec M, Crowell EF, Jouy H, Pochylova Z, Parcy F, Höfte H, Gonneau M, Vernhettes S (2007) Organization of cellulose synthase complexes involved in primary cell wall synthesis in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 104:15572–15577
Dhugga KS, Barreiro R, Whitten B, Stecca K, Hazebroek J, Randhawa GS, Dolan M, Kinney AJ, Tomes D, Nichols S, Anderson P (2004) Guar seed betamannan synthase is a member of the cellulose synthase super gene family. Science 303:363–366
Dill GM, Cajacob CA, Padgette SR (2008) Glyphosate-resistant crops: adoption, use and future considerations. Pest Manag Sci 64:326–331
Din N, Gilkes RN, Tekant B, Miller RC Jr, Warren RAJ, Kilburn DG (1991) Non-hydrolytic disruption of cellulose fibers by the binding domain of a bacterial cellulase. Bio/Technology 9:1096–1099
Doblin MS, Pettolino FA, Wilson SM, Campbell R, Burton RA, Fincher GB, Newbigin E, Bacic A (2009) A barley cellulose synthase-like CSLH gene mediates (1,3;1,4)-beta-D-glucan synthesis in transgenic Arabidopsis. Proc Natl Acad Sci 106:5996–6001
Doebley J, Stec A, Gustus C (1995) *teosinte branched1* and the origin of maize: evidence for epistasis and the evolution of dominance. Genetics 141:333–346
Doebley J, Stec A, Hubbard L (1997) The evolution of apical dominance in maize. Nature 386:485–488
Doi K, Izawa T, Fuse T, Yamanouchi U, Kubo T, Shimatani Z, Yano M, Yoshimura A (2004) Ehd1, a B-type response regulator in rice, confers short-day promotion of flowering and controls FT-like gene expression independently of Hd1. Genes Dev 18:926–936
Duan YX, Guo WW, Meng HJ, Tao NG, Li DD, Deng XX (2007) High efficient transgenic plant regeneration from embryogenic calluses of *Citrus sinensis*. Biol Plant 51:212–216
Ebringerová A, Heinze T (2000) Xylan and xylan derivatives – biopolymers with valuable properties. I. Naturally occurring xylans structures, isolation procedures and properties. Macromol Rapid Commun 21:542–556
Eichenseer H, Strohbehn R, Burks J (2008) Frequency and severity of western bean cutworm (Lepidoptera: Noctuidae) ear damage in transgenic corn hybrids expressing different *Bacillus thuringiensis* cry toxins. J Econ Entomol 101:555–563
Ellis RT, Stockhoff BA, Stamp L, Schnepf HE, Schwab GE, Knuth M, Russell J, Cardineau GA, Narva KE (2002) Novel *Bacillus thuringiensis* binary insecticidal crystal proteins active on western corn rootworm, *Diabrotica virgifera virgifera* LeConte. Appl Environ Microbiol 68:1137–1145

Engler DE, Chen X (2009) Transformation and engineered trait modification in *Miscanthus* species. WO/2009/132116

European Union (2012) Environmental releases of GMOs, European Union. http://mbg.jrc.ec.europa.eu/deliberate/dbplants.asp

Ercoli L, Mariotti M, Mason A, Bonari E (1999) Effect of irrigation and nitrogen fertilization on biomass yield and effciency of energy use in crop production of *Miscanthus*. Field Crop Res 63:3–11

Feltus FA, Hart GE, Schertz KF, Casa AM, Kresovich S, Abraham S, Klein PE, Brown PJ, Paterson AH (2006) Alignment of genetic maps and QTLs between inter- and intra-specific sorghum populations. Theor Appl Genet 112:1295–1305

Fernandez MGS, Becraft PW, Yin Y, Lübberstedt T (2009) From dwarves to giants? Plant height manipulation for biomass yield. Trends Plant Sci 14:454–461

Fincher GB (2009) Revolutionary times in our understanding of cell wall biosynthesis and remodeling in the grasses. Plant Physiol 149:27–37

Fowler S, Lee K, Onouchi H, Samach A, Richardson K, Morris B, Coupland G, Putterill J (1999) GIGANTEA: a circadian clock controlled gene that regulates photoperiodic flowering in *Arabidopsis* and encodes a protein with several possible membrane-spanning domains. EMBO J 18:4679–4688

Frame BR, Zhang H, Cocciolone SM, Sidorenko LV, Dietrich CR, Pegg SE, Zhen S, Schnable PS, Wang K (2000) Production of transgenic maize from bombarded Type II callus: effect of gold particle size and callus morphology on transformation efficiency. In Vitro Cell Dev Biol Plant 36:21–29

Franke R, Humphreys JM, Hemm MR, Denault JW, Ruegger MO, Cusumano JC, Chapple C (2002) The Arabidopsis REF8 gene encodes the 3-hydroxylase of phenylpropanoid metabolism. Plant J 30:33–45

Fujii S, Hayashi T, Mizuno K (2010) Sucrose synthase is an integral component of the cellulose synthesis machinery. Plant Cell Physiol 51:294–301

Fujioka S, Yamane H, Spray CR, Katsumit M, Phinney BO, Gaskin P, Macmillan J, Takahashiii N (1988) The dominant non-gibberellin-responding dwarf mutant (D8) of maize accumulates native gibberellins. Proc Natl Acad Sci USA 85:9031–9035

Gams W, Klamer M, O'Donnell K (1999) *Fusarium miscanthi* sp. nov. from *Miscanthus* litter. Mycology 91:263–268

Gardiner JC, Taylor NG, Turner SR (2003) Control of cellulose synthase complex localization in developing xylem. Plant Cell 8:1740–1748

Ghosh A, Ganapathi TR, Nath P, Bapat VA (2009) Establishment of embryogenic cell suspension cultures and *Agrobacterium*-mediated transformation in an important *Cavendish banana* cv. Robusta (AAA). Plant Cell Tissue Organ Culture 97:131–139

Girio FM, Fonseca C, Carvalheiro F, Duarte LC, Marques S, Bogel-Łukasik R (2010) Hemicelluloses for fuel ethanol: a review. Bioresour Technol 101:4775–4800

Glowacka K, Jezowski KZ (2010) The effects of genotype, inflorescence developmental stage and induction medium on callus induction and plant regeneration in two *Miscanthus* species. Plant Cell Tissue Organ Culture 102:79–86

Goicoechea M, Lacombe E, Legay S, Mihaljevic S, Rech P, Jauneau A, Lapierre C, Pollet B, Verhaegen D, Chaubet-Gigot N, Grima-Pettenati J (2005) EgMYB2, a new transcriptional activator from *Eucalyptus xylem*, regulates secondary cell wall formation and lignin biosynthesis. Plant J 43:553–567

Gonsalves D (1998) Resistance to papaya ringspot virus. Annu Rev Phytopathol 36:415–437

Gossmann M (2000) Schadwirkung einer pilzparasitären Rhizombesiedlung und Maßnahmen zur Verbesserung der Austriebs-und Biomasseleistung bei *Miscanthus* × *giganteus* Greef et Deu. In: Pude R (ed) Miscanthus-Vom Anbau bis zur Verwertung *Miscanthus*-symposium, Bonn, Beiträge Agrarwissenschaften, pp 26–31

Gravois KA, Milligan SB, Martin FA (1991) Indirect selection for increased sucrose yield in early sugarcane testing stages. Field Crop Res 26:67–73

Gray KA, Zhao L, Emptage M (2006) Bioethanol. Curr Opin Chem Biol 10:141–146

Greenberg SM, Adamczyk JJ Jr (2007) Noctuid survivorship and damage in Widestrike, Bollgard, and Bollgard II cottons in the lower Rio Grande valley of Texas. In: Dugger P, Richter D (eds) Proceedings of the beltwide cotton conference, New Orleans, LA, 9–12 January 2007. National Cotton Council, Memphis, TN, pp 316–320

Green JM, Hazel CB, Forney DR, Pugh LM (2008) New multiple-herbicide crop resistance and formulation technology to augment the utility of glyphosate. Pest Manag Scie 64:332–339

Greenup A, Peacock WJ, Dennis ES, Trevaskis B (2009) The molecular biology of seasonal flowering-responses in Arabidopsis and the cereals. Ann Bot 103:1165–1172

Gu YQ, Wildermuth MC, Chakravarthy S, Loh YT, Yang C, He X, Han Y, Martin GB (2002) Tomato transcription factors pti4, pti5, and pti6 activate defense responses when expressed in Arabidopsis. Plant Cell 14:817–831

Guiderdoni E, Demarly Y (1988) Histology of somatic embryogenesis in cultured leaf segments of sugarcane plantlets. Plant Cell Tissue Organ Culture 14:71–88

Gupta SC (1995) Inheritance and allelic study of brown midrib trait in pearl-millet. J Hered 86:301–303

Gutterson N, Reuber TL (2004) Regulation of disease resistance pathways by AP2/ERF transcription factors. Curr Opin Plant Biol 7:465–471

Halbert SE, Remaudiere G (2000) A new oriental *Melanaphis* species recently introduced in North America [Hemiptera, Aphididae]. Rev Fr Entomol 22:109–117

Halpin C, Holt K, Chojecki J, Oliver D, Chabbert B, Monties B, Edwards K, Foxon GA (1998) Maize brown-midrib (bm1) – a mutation affecting the cinnamyl alcohol dehydrogenase gene. Plant J 14:545–553

Haney LLH, Hake SSH, Scott MPS (2008) Allelism testing of Maize Coop Stock Center lines containing unknown brown midrib alleles. Maize Newsl 82:4–5

Harris D, Stork J, Debolt S (2009) Genetic modification in cellulose-synthase reduces crystallinity and improves biochemical conversion to fermentable sugar. GCB Bioenergy 1:51–61

Hayama R, Izawa T, Shimamoto K (2002) Isolation of rice genes possibly involved in the photoperiodic control of flowering by a fluorescent differential display method. Plant Cell Physiol 43:494–504

Hayama R, Yokoi S, Tamaki S, Yano M, Shimamoto K (2003) Adaptation of photoperiodic control pathways produces short day flowering in rice. Nature 422:719–722

Hayashi T, Yoshida K, Park YW, Konishi T, Baba K (2005) Cellulose metabolism in plants. Int Rev Cytol 247:1–34

Hazen SP, Scott-Craig JS, Walton JD (2002) Cellulose synthase-like genes of rice. Plant Physiol 128:336–340

Heaton E, Voigt T, Long SP (2004a) A quantitative review comparing the yields of two candidate C4 perennial biomass crops in relation to nitrogen, temperature and water. Biomass Bioener 27:21–30

Heaton EA, Clifton-Brown J, Voigt TB, Jones MB, Long SP (2004b) Miscanthus for renewable energy generation: European Union experience and projections for Illinois. Mitig Adapt Strat Global Chang 9:433–451

Hellens R, Mullineaux P (2000) A guide to *Agrobacterium* binary Ti vectors. Trends Plant Sci 5:446–451

Herbers K, Wilke I, Sonnewald U (1995) A thermostable xylanase from *Clostridium thermocellum* expressed at high levels in the apoplast of transgenic tobacco has no detrimental effects and is easily purified. Biotechnology 13:63–66

Higgins JA, Bailey PC, Laurie DA (2010) Comparative genomics of flowering time pathways using *Brachypodium distachyon* as a model for the temperate grasses. PLoS One 5:e10065

Hisano H, Nandakumar R, Wang Z-Y (2009) Genetic modification of lignin biosynthesis for improved biofuel production. In Vitro Cell Dev Biol Plant 45:306–313

Hodgson EM, Lister SJ, Bridgwater AV, Clifton-Brown J, Donnison IS (2010) Genotypic and environmentally derived variation in the cell wall composition of *Miscanthus* in relation to its use as a biomass feedstock. Biomass Bioener 34:652–660

Hoerlein G (1994) Glufosinate (phosphinothricin), a natural amino acid with unexpected herbicidal properties. Rev Environ Contam Toxicol 138:73–145

Holme IB, Petersen KK (1996) Callus induction and plant regeneration from different explant types of *Miscanthus* × *ogiformis* Honda 'Giganteus'. Plant Cell Tissue Organ Culture 45:43–52

Hu WJ, Harding SA, Lung J, Popko JL, Ralph J, Stokke DD, Tsai CJ, Chiang VL (1999) Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees. Nat Biotechnol 17:808–812

Hu FY, Tao DY, Sacks E, Fu BY, Xu P, Li J, Yang Y, McNally K, Khush KS, Paterson AH, Li Z-K (2003) Convergent evolution of perenniality in rice and sorghum. Proc Natl Acad Sci USA 100:40504054

Hu H, Dai M, Yao J, Xiao B, Li X, Zhang Q, Xiong L (2006) Overexpressing a NAM, ATAF, and CUC (NAC) transcription factor enhances drought resistance and salt tolerance in rice. Proc Natl Acad Sci USA 103:12987–12992

Hu H, You J, Fang Y, Zhu X, Qi Z, Xiong L (2008) Characterization of transcription factor gene SNAC2 conferring cold and salt tolerance in rice. Plant Mol Biol 67:169–181

Huggett DAJ, Leather SR, Walters KFA (1999) Suitability of the biomass crop *Miscanthus sinensis* as a host for the aphids *Rhopalosiphum padi* (L.) and *Rhopalosiphum maidis* (F.), and its susceptibility to the plant luteovirus barley yellow dwarf virus. Agric For Entomol 1:143–149

Hung K-H, Chiang T-Y, Chiu C-T, Hsu T-W, Ho C-W (2009) Isolation and characterization of microsatellite loci from a potential biofuel plant *Miscanthus sinensis* (Poaceae). Conserv Genet 10:1377–1380

Huntley SK, Ellis D, Gilbert M, Chapple C, Mansfield SD (2003) Significant increases in pulping efficiency in C4H:F5H-transformed poplars: improved chemical savings and reduced environmental toxins. J Agric Food Chem 51:6178–6183

ISAAA (2010) ISAAA brief 41-2010: executive summary (2010) global status of commercialized biotech/GM crops: 2009 the first fourteen years, 1996 to 2009

Ito Y, Katsura K, Maruyama K, Taji T, Kobayashi M, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2006) Functional analysis of rice DREB1/CBF-type transcription factors involved in cold-responsive gene expression in transgenic rice. Plant Cell Physiol 47:141–153

Itoh T, Kimura S (2001) Immunogold labeling of terminal cellulose synthesizing complexes. J Plant Res 114:483–489

Itoh H, Ueguchi-Tanaka M, Sentoku N, Kitano H, Matsuoka M, Kobayashi M (2001) Cloning and functional analysis of two gibberellin 3b-hydroxylase genes that are differently expressed during the growth of rice. Proc Natl Acad Sci USA 98:8909–8914

Izawa T, Takahashi Y, Yano M (2003) Comparative biology comes into bloom: genomic and genetic comparison of flowering pathways in rice and Arabidopsis. Curr Opin Plant Biol 6: 113–120

James VA, Neibaur I, Altpeter F (2008) Stress inducible expression of the DREB1A transcription factor from xeric, *Hordeum spontaneum* L. in turf and forage grass (*Paspalum notatum* Flugge) enhances abiotic stress tolerance. Transgenic Res 17:93–104

Jang JY, Choi C, Hwang HDJ (2010) The WRKY superfamily of rice transcription factors. Plant Pathol J 26:110–114

Jefferson RA (1989) The GUS reporter gene system. Nature 342:837–838

Jensen CS, Salchert K, Nielsen KK (2001) A TERMINAL FLOWER1-like gene from perennial ryegrass involved in floral transition and axillary meristem identity. Plant Physiol 125:1517–1528

Jensen CS, Salchert K, Gao C, Andersen C, Didion T, Nielsen KN (2004) Floral inhibition in red fescue (*Festuca rubra* L.) through expression of a heterologous flowering repressor from Lolium. Mol Breed 13:37–48

Jeoung JM, Krishnaveni S, Muthukrishnan S, Trick HN, Liang GH (2002) Optimization of sorghum transformation parameters using genes for green fluorescent protein and glucuronidase as visual markers. Hereditas 137:20–28

Jezowski S (2008) Yield traits of six clones of *Miscanthus* in the first 3 years following planting in Poland. Ind Crop Prod 27:65–68

Jiao Y, Wang Y, Xue D, Wang J, Yan M, Liu G, Dong G, Zeng D, Lu Z, Zhu X, Qian Q, Li J (2010) Regulation of *OsSPL14* by OsmiR156 defines ideal plant architecture in rice. Nat Genet 42:541–544

Jin J, Huang W, Gao JP, Yang J, Shi M, Zhu MZ, Luo D, Lin HX (2008) Genetic control of rice plant architecture under domestication. Nat Genet 40:1365–1369

Jing S, Zhou X, Song Y, Yu D (2009) Heterologous expression of *OsWRKY23* gene enhances pathogen defense and dark-induced leaf senescence in *Arabidopsis*. Plant Growth Regul 58:181–190

Johansson H, Sterky F, Amini B, Lundeberg J, Kleczkowski LA (2002) Molecular cloning and characterization of a cDNA encoding poplar UDP-glucose dehydrogenase, a key gene of hemicellulose/pectin formation. Biochim Biophys Acta 1576:53–58

Jorgenson LR (1931) Brown midrib in maize and its linkage relations. J Am Soc Agron 23:549–557

Kardailsky K, Shukla VK, Ahn JH, Dagenais N, Christensen SK, Nguyen JT, Chory J, Maria J, Harrison MJ, Weigel D (1999) Activation tagging of the floral inducer FT. Science 286: 1962–1965

Kasuga M, Miura S, Shinozaki K, Yamaguchi-Shinozaki K (2004) A combination of the Arabidopsis DREB1A gene and stress-inducible rd29A promoter improved drought- and low-temperature stress tolerance in tobacco by gene transfer. Plant Cell Physiol 45:346–350

Kausch AP, Hague J, Oliver MJ, Yi L, Daniell H, Mascia P, Watrud LS, Stewart N (2010) Transgenic Biofuel Feedstocks and Strategies for Biocontainment. Biofuels 1:163–176

Kavousi B, Daudi A, Cook CM, Joseleau JP, Ruel K, Devoto A, Bolwell GP, Blee KA (2010) Consequences of antisense down-regulation of a lignification-specific peroxidase on leaf and vascular tissue in tobacco lines demonstrating enhanced enzymic saccharification. Phytochemistry 71:531–542

Khan T, Reddy VS, Leelavathi S (2010) High-frequency regeneration via somatic embryogenesis of an elite recalcitrant cotton genotype (*Gossypium hirsutum* L.) and efficient *Agrobacterium*-mediated transformation. Plant Cell Tissue Organ Culture 101:323–330

Kim C, Zhang D, Auckland SA, Rainville LK, Jakob K, Kronmiller B, Sacks EJ, Deuter M, Paterson AH (2012) SSR-based genetic maps of *Miscanthus sinensis* and *M. sacchariflorus*, and their comparison to sorghum. TAG 124:1325–1338

Kimura S, Laosinchai W, Itoh T, Cui XJ, Linder CR, Brown RM (1999) Immunogold labeling of rosette terminal cellulose-synthesizing complexes in the vascular plant *Vigna angularis*. Plant Cell 11:2075–2085

Kimura T, Mizutani T, Tanaka T, Koyama T, Sakka K, Ohmiya K (2003) Molecular breeding of transgenic rice expressing a xylanase domain of the xynA gene from *Clostridium thermocellum*. Appl Microbiol Biotechnol 62:374–379

Kimura T, Mizutani T, Sun JL, Kawazu T, Karita S, Sakka M, Kobayashi Y, Ohmiya K, Sakka K (2010) Stable production of thermotolerant xylanase B of *Clostridium stercorarium* in transgenic tobacco and rice. Biosci Biotechnol Biochem 74:954–960

Kishore G, Shah D (1988) Amino acid biosynthesis inhibitors as herbicides. Annu Rev Biochem 57:627–663

Klein RR, Mullet JE, Jordan DR, Miller FR, Rooney WL, Menz MM, Franks CD, Klein PE (2008) The effect of tropical Sorghum conversion and inbred development on genome diversity as revealed by high-resolution genotyping. Plant Genome 48:S12–S26

Knoth C, Ringler J, Dangl JL, Eulgem T (2007) Arabidopsis WRKY70 is required for full RPP4-mediated disease resistance and basal defense against *Hyaloperonospora parasitica*. Mol Plant Microbe Interact 20:120–128

Kobayashi Y, Kaya H, Goto K, Iwabuchi M, Takashi Araki T (1999) A pair of related genes with antagonistic roles in mediating flowering signals. Science 286:1960–1962

Kojiama S, Takahashi Y, Kobayashi Y, Monna L, Sasaki T, Araki T, Yano M (2002) Hd3a, a rice ortholog of the Arabidopsis FT gene, promotes transition to flowering downstream of Hd1 under short-day conditions. Plant Cell Physiol 43:1096–1105

Komeda Y (2004) Genetic regulation of time to flower in *Arabidopsis thaliana*. Annu Rev Plant Biol 5:521–535

Koornneef M, Alonso-Blanco C, Peeters AJ, Soppe W (1998) Genetic control of flowering time in Arabidopsis. Annu Rev Plant Physiol Plant Mol Biol 49:345–370

Korban SS, Gasic K, Li X (2006) Rose (*Rosa hybrida* L.). Meth Mol Biol 344:351–358

Kozaki A, Hake S, Colasanti J (2004) The maize ID1 flowering time regulator is a zinc finger protein with novel DNA binding properties. Nucleic Acids Res 32:1710–1720

Kubo M, Udagawa M, Nishikubo N, Horiguchi G, Yamaguchi M, Ito J, Mimura T, Fukuda H, Demura T (2005) Transcription switches for protoxylem and metaxylem vessel formation. Genes Dev 19:1855–1860

Kuc J, Nelson OE (1964) The abnormal lignins produced by the brown-midrib mutants of maize. I. The brown-midrib-1 mutant. Arch Biochem Biophys 105:103–113

Kumria R, Waie B, Rajam MV (2001) Plant regeneration from transformed embryogenic callus of an elite indica rice via Agrobacterium. Plant Cell Tissue Organ Culture 67:63–71

Kunkel BN, Brooks DM (2002) Cross talk between signaling pathways in pathogen defense. Curr Opin Plant Biol 5:325–331

Lagaert S, Beliën T, Volckaert G (2009) Plant cell walls: protecting the barrier from degradation by microbial enzymes. Semin Cell Dev Biol 20:1064–1073

LaMondia JA (1996) Response of additional herbaceous perennial ornamentals to *Meloidogyne hapla*. J Nematol 28:636–638

Lechtenberg VL, Muller LD, Bauman LF, Rhykerd CL, Barnes RF (1972) Laboratory and in vitro evaluation of inbred and F2 populations of brown midrib mutants of *Zea mays* L. Agron J 64:657–660

Lee KY, Townsend J, Tepperman J, Black M, Chuil CF, Mazur B, Dunsmuir P, Bedbrook J (1988) The molecular basis of sulfonylurea herbicide resistance in tobacco. EMBO 7:1241–1248

Lee C, O'Neill MA, Tsumuraya Y, Darvill AG, Ye ZH (2007) The *irregular xylem9* mutant is deficient in xylan xylosyltransferase activity. Plant Cell Physiol 48:1624–1634

Lee C, Teng Q, Huang W, Zhong R, Ye ZH (2009a) The F8H glycosyltransferase is a functional paralog of FRA8 involved in glucuronoxylan biosynthesis in *Arabidopsis*. Plant Cell Physiol 50:812–827

Lee C, Teng Q, Huang W, Zhong R, Ye ZH (2009b) Down-regulation of PoGT47C expression in poplar results in a reduced glucuronoxylan content and an increased wood digestibility by cellulase. Plant Cell Physiol 50:1075–1089

Lee C, Teng Q, Huang W, Zhong R, Ye ZH (2009c) The poplar GT8E and GT8F glycosyltransferases are functional orthologs of Arabidopsis PARVUS involved in glucuronoxylan biosynthesis. Plant Cell Physiol 50:1982–1987

Lee C, Teng Q, Huang W, Zhong R, Ye ZH (2010) The Arabidopsis family GT43 glycosyltransferases form two functionally nonredundant groups essential for the elongation of glucuronoxylan backbone. Plant Physiol 153:526–541

Lee D, Natesan E (2006) Evaluating genetic containment strategies for transgenic plants. Trends in Biotech 24:109–114

Legay S, Lacombe E, Goicoechea M, Brière C, Séguin A, Mackay J, Grima-Pettenati J (2007) Molecular characterization of *EgMYB1*, a putative transcriptional repressor of the lignin biosynthetic pathway. Plant Sci 173:542–549

Lewis JM, Mackintosh CA, Shin S, Gilding E, Kravchenko S, Baldridge G, Zeyen R, Muehlbauer GJ (2008) Overexpression of the maize Teosinte Branched1 gene in wheat suppresses tiller development. Plant Cell Rep 27:1217–1225

Li X, Qian Q, Fu Z, Wang Y, Xiong G, Zeng D, Wang X, Liu X, Teng S, Hiroshi F, Yuan M, Luo D, Han B, Li J (2003) Control of tillering in rice. Nature 422:618–621

Li J, Brader G, Palva ET (2004) The WRKY70 transcription factor: a node of convergence for jasmonate-mediated and salicylate-mediated signals in plant defense. Plant Cell 16:319–331

Li D, Wang L, Wang M, Xu YY, Luo W, Liu YJ, Xu ZH, Li J, Chong K (2009) Engineering OsBAK1 gene as a molecular tool to improve rice architecture for high yield. Plant Biotechnol J 7:791–806

Liepman AH, Wilkerson CG, Keegstra K (2005) Expression of cellulose synthase-like (Csl) genes in insect cells reveals that CslA family members encode mannan synthases. Proc Natl Acad Sci USA 102:2221–2226

Lin Y, Tanaka S (2006) Ethanol fermentation from biomass resources: current state and prospects. Appl Microbiol Biotechnol 69:627–642

Lin YR, Schertz KF, Paterson AH (1995) Comparative analysis of QTLs affecting plant height and maturity across the Poaceae, in reference to an interspecific sorghum population. Genetics 141:391–411

Long SP, Zhu XG, Naidu SL, Ort DR (2006) Can improvement in photosynthesis increase crop yields? Plant Cell Environ 29:315–330

Longo C, Lickwar C, Hu Q, Nelson-Vasilchik K, Viola D, Hague J, Chandlee JM, Luo H, Kausch AP (2006) Turf grasses. Methods Mol Biol 344:83–95

Lopez-Perez AJ, Carreño J, Dabauza M (2009) Transformation of embryogenic callus and transgenic plant regeneration in table grapevine 'sugraone' (*Vitis vinifera* L.): effect of *Agrobacterium tumefaciens* strain. Acta Hort 827:415–420

Lu C, Vasil IK, Ozias-Akins P (1982) Somatic embryogenesis in *Zea mays* L. Theor Appl Genet 62:109–112

Lübberstedt Th, Melchinger AE, Schön CC, Utz HF, Klein D (1997) QTL mapping in testcrosses of European flint lines of maize: I. Comparison of different testers for forage yield traits. Crop Sci 37:921–931

Luo A, Qian Q, Yin H, Liu X, Yin C, Lan Y, Tang J, Tang Z, Cao S, Wang X, Xia K, Fu X, Luo D, Chu C (2006) EUI1, encoding a putative cytochrome P450 monooxygenase, regulates internode elongation by modulating gibberellin responses in rice. Plant Cell Physiol 47:181–191

Lynd LR, Laser MS, Bransby D, Dale BE, Davison B, Hamilton R, Himmel M, Keller M, McMillan JM, Sheehan J, Wyman CE (2008) How biotech can transform biofuels. Nat Biotechnol 26:169–172

Ma X-F, Jensen E, Alexandrov N, Troukhan M, Zhang L, Thomas-Jones S, Farrar K, Clifton-Brown J, Donnison I, Flavell R (2012) High Resolution Genetic Mapping by Genome Sequencing Reveals Genome PLoS ONE 7(3): e33821. doi:10.1371/journal.pone.0033821

McCarthy RL, Zhong R, Ye ZH (2009) MYB83 is a direct target of SND1 and acts redundantly with MYB46 in the regulation of secondary cell wall biosynthesis in Arabidopsis. Plant Cell Physiol 50:1950–1964

Melzer S, Lens F, Gennen J, Vanneste S, Rohde A, Beeckman T (2008) Flowering-time genes modulate meristem determinacy and growth form in *Arabidopsis thaliana*. Nat Genet 40:1489–1492

Menden B, Kohlhoff M, Moerschbacher BM (2007) Wheat cells accumulate a syringyl-rich lignin during the hypersensitive resistance response. Phytochem 68:513–520

Merkele SA, Parrott W, Flin BS (1995) Morphogenic aspect of somatic embryogenesis. In: Thorpe TA (ed) In vitro embryogenesis in plants. Kluwer Academic Publishers, Dordrecht, pp 155–203

Meyer K, Shirley AM, Cusumano JC, Bell-Lelong DA, Chapple C (1998) Lignin monomer composition is determined by the expression of a cytochrome P450-dependent monooxygenase in Arabidopsis. Proc Natl Acad Sci USA 95:6619–6623

Michaels SD, Amasino RM (1999) FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. Plant Cell 11:949–956

Miller JD, Tai PY, Edme SJ, Comstock JC, Glaz BS, Gilbert RA (2005) Basic germplasm utilization in the sugarcane development program at Canal Point, FL, USA. Int Soc Sugar Cane Technol Proc 2:532–536

Miller TA, Muslin EH, Dorweiler JE (2008) A maize CONSTANS-like gene, conz1, exhibits distinct diurnal expression patterns in varied photoperiods. Planta 227:1377–1388

Ming R, Del Monte TA, Hernandez E, Moore PH, Irvine JE, Paterson AH (2002) Comparative analysis of QTLs affecting plant height and flowering among closely-related diploid and polyploid genomes. Genome 45:794–803

Mitsuda N, Iwase A, Yamamoto H, Yoshida M, Seki M, Shinozaki K, Ohme-Takagi M (2007) NAC transcription factors, NST1 and NST3, are key regulators of the formation of secondary walls in woody tissues of *Arabidopsis*. Plant Cell 19:270–280

Miura K, Ikeda M, Matsubara A, Song XJ, Ito M, Asano K, Matsuoka M, Kitano H, Ashikari M (2010) *OsSPL14* promotes panicle branching and higher grain productivity in rice. Nat Genet 42:545–549

Mizoguchi T, Wright L, Fujiwara S, Cremer F, Lee K, Onouchi H, Mouradov A, Fowler S, Kamada H, Putterill J, Coupland G (2005) Distinct roles of GIGANTEA in promoting flowering and regulating circadian rhythms in *Arabidopsis*. Plant Cell 17:2255–2270

Mullet JE, Rooney WL, Klein PE, Morishige D, Murphy R, Brady JA (2010) Discovery and utilization of sorghum genes (MA5/MA6). Patent Application Number: 20100024065

Murray SC, Rooney WL, Mitchell SE, Sharma A, Klein PE, Mullet JE, Kresovich S (2008) Genetic improvement of sorghum as a biofuel feedstock: II. QTL for stem and leaf structural carbohydrates. Crop Sci 48:2180–2193

Muszynski MG, Dam T, Li B, Shirbroun DM, Hou Z, Bruggemann E, Archibald R, Ananiev EV, Danilevskaya ON (2006) Delayed flowering1 Encodes a basic leucine zipper protein that mediates floral inductive signals at the shoot apex in maize. Plant Physiol 142:1523–1536

Nakagawa M, Shimamoto K, Kyozuka J (2002) Overexpression of RCN1 and RCN2, rice TERMINAL FLOWER 1/CENTRORADIALIS homologs, confers delay of phase transition and altered panicle morphology in rice. Plant J 29:743–750

Neuffer MG, Coe EH, Wessler SR (1997) Mutants of maize. Cold Spring Harbor Laboratory Press, Plainview, New York

Nida DL, Kolacz KH, Buehler RE, Deaton WR, Schuler WR, Armstrong TA, Taylor ML, Ebert CC, Rogan GJ (1996) Glyphosate-tolerant cotton: genetic characterization and protein expression. J Agric Food Chem 44:1960–1966

O'Neill NR, Farr DF (1996) Miscanthus blight; a new foliar disease of ornamental grasses and sugarcane incited by *Leptosphaeria* sp. and its anamorphic state *Stagonospora* sp. Plant Dis 80:980–987

Oomen RJFJ, Doeswijk-Voragen CHL, Bush MS, Vincken JP, Borkhardt B, Van de Broek LAM, Corsar J, Ulvskov P, Voragen AGJ, McCann MC, Visser RGF (2002) *In muro* fragmentation of the rhamnogalacturonan I backbone in potato (*Solanum tuberosum* L.) results in a reduction and altered location of the galactan and arabinan side-chains and abnormal periderm development. Plant J 30:403–413

Padgette SR, Kolacz KH, Delannay X, Re DB, La Vallee BJ, Tinius CN, Rhodes WK, Otero YI, Barry GF, Eichholtz DA, Peschke VM, Nida DL, Taylor NB, Kishore GM (1995) Development, identification, and characterization of a glyphosate-tolerant soybean line. Crop Sci 35:1451–1461

Pandey SP, Somssich IE (2009) The role of WRKY transcription factors in plant immunity. Plant Physiol 150:1648–1655

Paredez AR, Somerville CR, Ehrhardt DW (2006) Visualization of cellulose synthase demonstrates functional association with microtubules. Sci 312:1491–1495

Park JM, Park CJ, Lee SB, Ham BK, Shin R, Paek KH (2001) Overexpression of the tobacco Tsi1 gene encoding an EREBP/AP2-type transcription factor enhances resistance against pathogen attack and osmotic stress in tobacco. Plant Cell 13:1035–1046

Park YW, Baba K, Furuta Y, Iida I, Sameshima K, Arai M, Hayashi T (2004) Enhancement of growth and cellulose accumulation by overexpression of xyloglucanase in poplar. FEBS Lett 564:183–187

Passas H, Poethig RS (1993) Vegetative and reproductive development in leafy1 and early flowering plants. MNL 67:91–92

Patel M, Johnson J, Brettell R, Jacobsen J, Xue G (2000) Transgenic barley expressing a fungal xylanase gene in the endosperm of the developing grains. Mol Breed 6:113–124

Paterson AH, Schertz KF, Lin YR, Liu SC, Chang YL (1995) The weediness of wild plants: molecular analysis of genes influencing dispersal and persistence of Johnsongrass, *Sorghum halepense* (L.) Pers. Proc Natl Acad Sci USA 92:6127–6131

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle

AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Rahman M, Ware D, Westhoff P, Mayer KFX, Messing J, Daniel S, Rokhsar DS (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556

Patzlaff A, McInnis S, Courtenay A, Surman C, Newman LJ, Smith C, Bevan MW, Mansfield S, Whetten RW, Sederoff RR, Campbell MM (2003) Characterization of a pine MYB that regulates lignification. Plant J 36:743–754

Pedersen JF, Vogel KP, Funnell DL (2005) Impact of reduced lignin on plant fitness. Crop Sci 45:812–819

Peña MJ, Zhong R, Zhou GK, Richardson EA, O'Neill MA, Darvill AG, York WS, Ye ZH (2007) Arabidopsis irregular xylem8 and irregular xylem9: implications for the complexity of glucuronoxylan biosynthesis. Plant Cell 19:549–563

Peng J, Richards DE, Hartley NM, Murphy GP, Devos KM, Flintham JE, Beales J, Fish LJ, Worland AJ, Pelica F, Sudhakar D, Christou P, Snape JW, Gale MD, Harberd NP (1999) 'Green revolution' genes encode mutant gibberellin response modulators. Nature 400:256–261

Penning BW, Hunter CT III, Tayengwa R, Eveland AL, Dugard CK, Olek AT, Vermerris W, Koch KE, McCarty DR, Davis MF, Thomas SR, McCann MC, Carpita NC (2009) Genetic resources for maize cell wall biology. Plant Physiol 151:1703–1728

Pereira MG, Lee M (1995) Identification of genomic regions affecting plant height in sorghum and maize. Theor Appl Genet 90:380–388

Persson S, Paredez A, Carroll A, Palsdottir H, Doblin M, Poindexter P, Khitrov N, Auer M, Somerville CR (2007) Genetic evidence for three unique components in primary cell-wall cellulose synthase complexes in Arabidopsis. Proc Natl Acad Sci USA 104:15566–15571

Porter KS, Axtell JD, Lechtenberg VL, Colenbrander VF (1978) Phenotype, fiber composition, and in vitro dry matter disappearance of chemically induced brown midrib (bmr) mutants of sorghum. Crop Sci 18:205–208

Pude R (2005) Bedeutung morphologischer, chemischer and physiologischer parameter sowie ihre interaktion zur beurteilung der baustofferzeugung unterschiedlicher Miscanthus-Herkuenfte. Beitraege zu Agrarwissenschaften 30, Publisher: Verlag P. Wehle, Hauptstrasse 144a, D-53474 Bad Neuenahr-Ahrweiler, ISBN 3-935307-30-6

Pude R, Diepenbrock W, Franken H, Greef JM (1996) Impact and causes of winter kills of *Miscanthus*. Mitteil Gesell Pflanz (Germany) 9:61–62

Qin F, Sakuma Y, Li J, Liu Q, Li YQ, Shinozaki K, Yamaguchi-Shinozaki K (2004) Cloning and functional analysis of a novel DREB1/CBF transcription factor involved in cold-responsive gene expression in *Zea mays* L. Plant Cell Physiol 45:1042–1052

Qiu Y, Yu D (2009) Over-expression of the stress-induced *OsWRKY45* enhances disease resistance and drought tolerance in *Arabidopsis*. Environ Exp Bot 65:35–47

Qiu D, Xiao J, Ding X, Xiong M, Cai M, Cao Y, Li X, Xu C, Wang S (2007) OsWRKY13 mediates rice disease resistance by regulating defense-related genes in salicylate- and jasmonate-dependent signaling. Mol Plant Microbe Interact 20:492–499

Quinby JR (1966) Fourth maturity locus in sorghum. Crop Sci 6:516–518

Quinby JR (1974) Sorghum improvement and the genetics of growth. Texas A&M University Press, College Station, TX

Quinby JR, Karper R (1945) The inheritance of three genes that influence time of floral initiation and maturity date in milo. J Am Soc Agron 37:916–936

Quinby JR, Karper RE (1954) Inheritance of height in sorghum. Agron J 46:211–216

Ransom C, Balan V, Biswas G, Dale B, Crockett E, Sticklen M (2007) Heterologous *Acidothermus cellulolyticus* 1,4-beta-endoglucanase E1 produced within the corn biomass converts corn stover into glucose. Appl Biochem Biotechnol 137–140:207–219

Ratcliffe OJ, Amaya I, Vincent CA, Rothstein S, Carpenter R, Coen ES, Bradley DJ (1998) A common mechanism controls the life cycle and architecture of plants. Development 125:1609–1615

Ratcliffe OJ, Bradley DJ, Coen ES (1999) Separation of shoot and floral identity in Arabidopsis. Development 126:1109–1120

Reddy MS, Chen F, Shadle G, Jackson L, Aljoe H, Dixon RA (2005) Targeted down-regulation of cytochrome P450 enzymes for forage quality improvement in alfalfa (*Medicago sativa* L.). Proc Natl Acad Sci USA 102:16573–16578

Register JC, Nelson RS (1992) Early events in plant virus infections: relationships with genetically engineered protection and host gene resistance. Semin Virol 3:441–451

Richmond TA, Somerville CR (2000) The cellulose synthase superfamily. Plant Physiol 124: 495–498

Ritter KB, Jordan DR, Chapman SC, Godwin ID, Mace ES, McIntyre CL (2008) Identification of QTL for sugar-related traits in a sweet × grain sorghum (*Sorghum bicolor* L. Moench) recombinant inbred population. Mol Breed 22:367–384

Robert S, Mouille G, Höfte H (2004) The mechanism and regulation of cellulose synthesis in primary walls: lessons from cellulose-deficient *Arabidopsis* mutants. Cellulose 11:351–364

Rooney WL, Aydin S (1999) Genetic control of a photoperiod-sensitive response in *Sorghum bicolor* (L.) Moench. Crop Sci 39:397–400

Ross CA, Liu Y, Shen QJ (2007) The WRKY gene family in rice (*Oryza sativa*). J Integr Plant Biol 49:827–842

Ruan YL, Llewellyn DJ, Furbank RT (2003) Suppression of sucrose synthase gene expression represses cotton fibre cell initiation, elongation, and seed development. Plant Cell 15:952–964

Saballos A, Vermerris W, Rivera L, Ejeta G (2008) Allelic association, chemical characterization and saccharification properties of brown midrib mutants of sorghum (*Sorghum bicolor* (L.) Moench). Bioenergy Res 1:193–204

Salvi S, Sponza G, Morgante M, Tomes D, Niu X, Fengler KA, Meeley R, Ananiev EV, Svitashev S, Bruggemann E, Li B, Hainey CF, Radovic S, Zaina G, Rafalski JA, Tingey SV, Miao GH, Phillips RL, Tuberosa R (2007) Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. Proc Natl Acad Sci USA 104:11376–11381

Sasaki A, Ashikari M, Ueguchi-Tanaka M, Itoh H, Nishimura A, Swapan D, Ishiyama K, Saito T, Kobayashi M, Khush GS, Kitano H, Matsuoka M (2002) Green revolution: a mutant gibberellin-synthesis gene in rice. Nature 416:701–702

Sasaki A, Itoh H, Gomi K, Ueguchi-Tanaka M, Ishiyama K, Kobayashi M, Jeong DH, An G, Kitano H, Ashikari M, Matsuoka M (2003) Accumulation of phosphorylated repressor for gibberellin signaling in an F-box mutant. Science 299:1896–1898

Sattler SE, Funnell-Harris DL, Pedersen JF (2010) Brown midrib mutations and their importance to the utilization of maize, sorghum, and pearl millet lignocellulosic tissues. Plant Sci 178:229–238

Scheible WR, Eshed R, Richmond T, Delmer D, Somerville CR (2001) Modifications of cellulose synthase confer resistance to isoxaben and thiazolidinone herbicides in *Arabidopsis* Ixr1 mutants. Proc Natl Acad Sci USA 98:10079–10084

Scheller HV, Ulvskov P (2010) Hemicelluloses. Annu Rev Plant Biol 61:263–289

Searle I, He Y, Turck F, Vincent C, Fornara F, Kröber S, Amasino RA, Coupland G (2006) The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling in Arabidopsis. Genes Dev 20:898–912

Sharabi-Schwager M, Lers A, Samach A, Guy CL, Porat R (2010) Overexpression of the CBF2 transcriptional activator in Arabidopsis delays leaf senescence and extends plant longevity. J Exp Bot 61:261–273

Sheldon CC, Burn JE, Perez PP, Metzger J, Edwards JA, Peacock WJ, Dennis ES (1999) The FLF MADS box gene: a repressor of flowering in Arabidopsis regulated by vernalization and methylation. Plant Cell 11:445–458

Sheldon CC, Rouse DT, Finnegan EJ, Peacock WJ, Dennis ES (2000) The molecular basis of vernalization: the central role of FLOWERING LOCUS C (FLC). Proc Natl Acad Sci USA 28(97):3753–3758

Shen QH, Saijo Y, Mauch S, Biskup C, Bieri S, Keller B, Seki H, Ülker B, Somssich IE, Schulze-Lefert P (2007) Nuclear activity of MLA immune receptors links isolate-specific and basal disease resistance responses. Science 315:1098–1103

Shinozaki K, Yamaguchi-Shinozaki K (2007) Gene networks involved in drought stress response and tolerance. J Exp Bot 58:221–227

Shoseyov O, Shani Z, Levy I (2006) Carbohydrate binding modules: biochemical properties and novel applications. Biol Rev 70:283–295

Sibout R, Baucher M, Gatineau M, van Doorsselaere J, Mila I, Pollet B, Maba B, Pilate G, Lapierre C, Boerjan W, Jouanin L (2002) Expression of a poplar cDNA encoding a ferulate-5-hydroxylase/coniferaldehyde 5-hydroxylase increases S lignin deposition in *Arabidopsis thaliana*. Plant Physiol Biochem 40:1087–1096

Siebert MW, Tindall KV, Leonard BR, van Duyn JW, Babcock JM (2008) Evaluation of corn hybrids expressing Cry1F (Herculex I insect protection) against fall armyworm (Lepidoptera: Noctuidae) in the southern United States. J Entomol Sci 43:41–51

Simmons BA, Loqué D, Ralph J (2010) Advances in modifying lignin for enhanced biofuel production. Curr Opin Plant Biol 13:1–8

Somerville C (2006) Cellulose synthesis in higher plants. Annu Rev Cell Dev Biol 22:53–78

Song Y, You J, Xiong L (2009) Characterization of OsIAA1 gene, a member of rice Aux/IAA family involved in auxin and brassinosteroid hormone responses and plant morphogenesis. Plant Mol Biol 70:297–309

Songstad DD, Petersen WL, Armstrong CL (1992) Establishment of friable embryogenic (type II) callus from immature tassels of *Zea mays* (Poaceae). Am J Bot 79:761–764

Soomro AF, Junejo S, Ahmed A, Aslam M (2006) Evaluation of different promising sugarcane varieties for some quantitative and qualitative attributes under thatta (Pakistan) conditions. Int J Agric Biol 8:195–197

Sørensen SO, Pauly M, Bush MS, Skjot M, McCann MC, Borkhardt B, Ulvskov P (2000) Pectin engineering: modification of potato pectin by *in vivo* expression of an endo-1,4-β-D-galactanase. Proc Natl Acad Sci USA 97:7639–7644

Spencer JL, Raghu S (2009) Refuge or reservoir? The potential impacts of the biofuel crop *Miscanthus*×*giganteus* on a major pest of maize. PLoS One 4(12):e8336. doi:10.1371

Sprague SJ, Marcroft SJ, Hayden HL, Howlett BJ (2006) Major gene resistance to blackleg in *Brassica napus* overcome within three years of commercial production in southeastern Australia. Plant Dis 90:190–198

Stewart SD, Adamczyk JJ Jr, Knighten KS, Davis FM (2001) Impact of Bt cottons expressing one or two insecticidal proteins of *Bacillus thuringiensis* Berliner on growth and survival on noctuid (Lepidoptera) larvae. J Econ Entomol 94:752–760

Stewart JJ, Akiyama T, Chapple C, Ralph J, Mansfield SD (2009) The effects on lignin structure of overexpression of ferulate 5-hydroxylase in hybrid poplar. Plant Phys 150:621–635

Sun F, Zhang W, Xiong G, Yan M, Qian Q, Li J, Wang Y (2010) Identification and functional analysis of the MOC1 interacting protein 1. J Genet Genome 37:69–77

Swaminathan K, Alabady MS, Varala K, de Paoli E, Ho I, Rokhsar DS, Arumuganathan AK, Ming R, Green PJ, Meyers BC, Moose SP, Hudson ME (2010) Genomic and small RNA sequencing of *Miscanthus*×*giganteus* shows the utility of sorghum as a reference genome sequence for Andropogoneae grasses. Genome Biol 11:R12

Taheri P, Tarighi S (2010) Riboflavin induces resistance in rice against *Rhizoctonia solani* via jasmonate-mediated priming of phenylpropanoid pathway. J Plant Physiol 167:201–208

Taylor NG (2008) Cellulose biosynthesis and deposition in higher plants. New Phytol 178:239–252

Taylor NG, Scheible WR, Cutler S, Somerville CR, Turner SR (1999) The irregular xylem3 locus of Arabidopsis encodes a cellulose synthase required for secondary cell wall synthesis. Plant Cell 11:769–780

Taylor NG, Laurie S, Turner SR (2000) Multiple cellulose synthase catalytic subunits are required for cellulose synthesis in Arabidopsis. Plant Cell 12:2529–2540

Taylor NG, Howells RM, Huttly AK, Vickers K, Turner SR (2003) Interactions among three distinct CesA proteins essential for cellulose synthesis. Proc Natl Acad Sci USA 100:1450–1455

Teakle NS, Shukla DD, Ford RE (1989) Sugarcane mosaic virus. AAB Descriptions of Plant Viruses, no. 342 (no. 88 revised)

Tenhaken R, Thulke O (1996) Cloning of an enzyme that synthesizes a key nucleotide-sugar precursor of hemicellulose biosynthesis from soybean: UDP-glucose dehydrogenase. Plant Physiol 112:1127–1134

Thinggaard K (1997) Study of the role of *Fusarium* in the field establishment problem of *Miscanthus*. Acta Agric Scand B Plant Soil Sci 47:238–241

Tiwari SB, Shen Y, Chang HC, Hou Y, Harris A, Ma SF, McPartland M, Hymus GJ, Adam L, Marion C, Belachew A, Repetti PP, Reuber TL, Ratcliffe OJ (2010) The flowering time regulator CONSTANS is recruited to the FLOWERING LOCUS T promoter via a unique cis-element. New Phytol 187:57–66

Tong JP, Liu XJ, Zhang SY, Li SQ, Peng XJ, Yang J, Zhu YG (2007) Identification, genetic characterization, GA response and molecular mapping of Sdt97: a dominant mutant gene conferring semi-dwarfism in rice (*Oryza sativa* L.). Genet Res 89:221–230

Trick HN, Finer JJ (1998) Sonication-assisted *Agrobacterium*-mediated transformation of soybean [*Glycine max* (L.) Merrill] embryogenic suspension culture tissue. Plant Cell Rep 17:482–488

Turck F, Fornara F, Coupland G (2008) Regulation and identity of florigen: FLOWERING LOCUS T moves center stage. Annu Rev Plant Biol 59:573–594

Turner SR, Somerville CR (1997) Collapsed xylem phenotype of Arabidopsis identifies mutants deficient in cellulose deposition in the secondary cell wall. Plant Cell 9:689–701

Ueguchi-Tanaka M, Nakajima M, Katoh E, Ohmiya H, Asano K, Saji S, Hongyu X, Ashikari M, Kitano H, Yamaguchi I, Matsuoka M (2007) Molecular interactions of a soluble gibberellin receptor, GID1, with a rice DELLA protein, SLR1, and gibberellins. Plant Cell 19:2140–2155

Uelker B, Mukhtar SM, Somssich IE (2007) The WRKY70 transcription factor of Arabidopsis influences both the plant senescence and defense signalling pathways. Planta 226:125–137

USDA-APHIS (2012) Status of permits, notifications, and petitions. United States Department of Agriculture, Animal and Plant Health Inspection Service. http://www.aphis.usda.gov/biotechnology/status.shtml

Vignols F, Rigau J, Torres MA, Capellades M, Puigdomenech P (1995) The *brown midrib* (*bm3*) mutation in maize occurs in the gene encoding caffeic acid *O*-methyltransferase. Plant Cell 7:407–416

Vogel J (2008) Unique aspects of the grass cell wall. Curr Opin Plant Biol 11:301–307

Wang ZY, Ge Y (2006) Recent advances in genetic transformation of forage and turf grasses. In Vitro Cell Dev Biol Plant 42:1–18

Wang Y, Li J (2006) Genes controlling plant architecture. Curr Opin Biotechnol 17:123–129

Wang RL, Stec A, Hey J, Lukens L, Doebley J (1999) The limits of selection during maize domestication. Nature 398:236–239

Wang ZZ, Zhang SZ, Yang BP, Li YR (2005) Trehalose synthase gene transfer mediated by *Agrobacterium tumefaciens* enhances resistance to osmotic stress in sugarcane. Sugar Tech 7:49–54

Wang H, Hao J, Chen X, Hao Z, Wang X, Lou U, Peng Y, Guo Z (2007) Overexpression of rice WRKY89 enhances ultraviolet B tolerance and disease resistance in rice plants. Plant Mol Biol 65:799–815

Wang L, Xu Y, Zhang C, Ma Q, Joo SH, Kim SK, Xu Z, Chong K (2008a) OsLIC, a novel CCCH-type zinc finger protein with transcription activation, mediates rice architecture via brassinosteroids signaling. PLoS One 3:e3521

Wang D, Portis AR Jr, Moose SP, Long SP (2008b) Cool C4 photosynthesis: pyruvate Pi dikinase expression and activity corresponds to the exceptional cold tolerance of carbon assimilation in *Miscanthus*×*giganteu*. Plant Physiol 148:557–567

Wang Q, Guan Y, Wu Y, Chen H, Chen F, Chu C (2008c) Overexpression of a rice OsDREB1F gene increases salt, drought, and low temperature tolerance in both Arabidopsis and rice. Plant Mol Biol 67:589–602

Wei X, Xu J, Guo H, Jiang L, Chen S, Yu C, Zhou Z, Hu P, Zhai H, Wan J (2010) DTH8 suppresses flowering in rice, influencing plant height and yield potential simultaneously. Plant Physiol 153:1747–1758

Welter ME, Clayton DS, Miller MA, Petolino JF (1995) Morphotypes of friable embryogenic maize callus. Plant Cell Rep 14:725–729

Wigge PA, Kim MC, Jaeger KE, Busch W, Schmid M, Lohmann JU, Weigel D (2005) Integration of spatial and temporal information during floral induction in *Arabidopsis*. Science 309:1056–1059

Wilczek AM, Roe JL, Knapp MC, Cooper MD, Lopez-Gallego C, Martin LJ, Muir CD, Sim S, Walker A, Anderson J, Egan F, Moyers BT, Petipas R, Giakountis A, Charbit E, Coupland G, Welch SM, Schmitt J (2009) Effects of genetic perturbation on seasonal life history plasticity. Science 323:930–934

Willrich MM, Braxton LB, Richburg JS, Lassiter RB, Langston VB, Haygood RA, Richardson JM, Hails FJ, Huckaba RM, Pellow JW, Thompson GD, Mueller JP (2005) Field and laboratory performance of WideStrike insect protection against secondary lepidopteran pests. In: Proceedings of the 2005 beltwide cotton conference, New Orleans, LA, 4–7 January 2005, pp 1262–1268

Wilmink A, Dons JJM (1993) Selective agents and marker genes for use in transformation of monocotyledonous plants. Plant Mol Biol Rep 11:165–185

Windels CE (2000) Economic and social impacts of *Fusarium* head blight: changing farms and rural communities in the Northern Great Plains. Phytopathology 90:17–21

Wu C, You C, Li C, Long T, Chen G, Byrne ME, Zhang Q (2008) RID1, encoding a Cys2/His2-type zinc finger transcription factor, acts as a master switch from vegetative to floral development in rice. Proc Natl Acad Sci USA 105:12915–12920

Wu AM, Hörnblad E, Voxeur A, Gerber L, Rihouey C, Lerouge P, Marchant A (2010) Analysis of the Arabidopsis IRX9/IRX9-L and IRX14/IRX14-L pairs of glycosyltransferase genes reveals critical contributions to biosynthesis of the hemicellulose glucuronoxylan. Plant Physiol 153:542–554

Wyman CE (2007) What is (and is not) vital to advancing cellulosic ethanol. Trends Biotechnol 25:153–157

Xi Y, Ge Y, Wang ZY (2009) Genetic transformation of switchgrass. Methods Mol Biol 581:53–59

Xie K, Wu C, Xiong L (2006) Genomic organization, differential expression, and interaction of SQUAMOSA promoter-binding-like transcription factors and microRNA156 in rice. Plant Physiol 142:280–293

Xin Z, Wang M, Burow G, Burke J (2009) An induced sorghum mutant population suitable for bioenergy research. Bioenergy Res 2:10–16

Xu ZS, Xia LQ, Chen M, Cheng XG, Zhang RY, Li LC, Zhao YX, Lu Y, Ni ZY, Liu L, Qiu ZG, Ma YZ (2007) Isolation and molecular characterization of the *Triticum aestivum* L. ethylene-responsive factor 1 (TaERF1) that increases multiple stress tolerance. Plant Mol Biol 65:719–732

Xu Z, Zhang D, Hu J, Zhou X, Ye X, Reichel KL, Stewart NR, Syrenne RD, Yang X, Gao P, Shi W, Doeppke C, Sykes RW, Burris JN, Bozell JJ, Cheng MZ, Hayes DG, Labbe N, Davis M, Stewart CN Jr, Yuan JS (2009) Comparative genome analysis of lignin biosynthesis gene families across the plant kingdom. BMC Bioinformatics 10(Suppl 11):S3

Yamaguchi M, Kubo M, Fukuda H, Demura T (2008) Vascular-related NAC-DOMAIN7 is involved in the differentiation of all types of xylem vessels in Arabidopsis roots and shoots. Plant J 55:652–664

Yamaguchi M, Goué N, Igarashi H, Ohtani M, Nakano Y, Mortimer JC, Nishikubo N, Kubo M, Katayama Y, Kakegawa K, Dupree P, Demura T (2010) Vascular-related NAC-DOMAIN6 and vascular-related NAC-DOMAIN7 effectively induce transdifferentiation into xylem vessel elements under control of an induction system. Plant Physiol 153:906–914

Yamashita S, Nonaka N, Namba S, Doi Y, Yora K (1985) Miscanthus streak virus, a geminivirus in *Miscanthus sacchariflorus* Benth. et Hook. Annu Phytopathol Soc Jpn 51:582–590

Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J (2003) Positional cloning of the wheat vernalization gene VRN1. Proc Natl Acad Sci USA 100:6263–6268

Yan L, Loukoianov A, Blechl A, Tranquilli G, Ramakrishna W, SanMiguel P, Bennetzen JL, Echenique V, Dubcovsky J (2004) The wheat VRN2 gene is a flowering repressor down-regulated by vernalization. Science 303:1640–1644

Yang A, He C, Zhang K, Zhang J (2006) Improvement of *Agrobacterium*-mediated transformation of embryogenic calluses from maize elite inbred lines. In Vitro Cell Dev Biol Plant 42: 215–219

Yang P, Wang Y, Bai Y, Meng K, Luo H, Yuan T, Fan Y, Yao B (2007) Expression of xylanase with high specific activity from *Streptomyces olivaceoviridis* A1 in transgenic potato plants (*Solanum tuberosum* L.). Biotechnol Lett 29:659–667

Yano M, Sasaki T (1997) Genetic and molecular dissection of quantitative traits in rice. Plant Mol Biol 35:145–153

Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, Baba T, Yamamoto K, Umehara Y, Nagamura Y, Sasaki T (2002) Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene CONSTANS. Plant Cell 12:2473–2484

Yanovsky MJ, Kay SA (2002) Molecular basis of seasonal time measurement in *Arabidopsis*. Nature 419:308–312

Yin Y, Huang J, Xu Y (2009) The cellulose synthase superfamily in fully sequenced plants and Algae. BMC Plant Biol 9:99

Zhong R, Ye ZH (2007) Regulation of cell wall biosynthesis. Curr Opin Plant Biol 10:564–572

Zhong R, Ye ZH (2010) The poplar PtrWNDs are transcriptional activators of secondary cell wall biosynthesis. Plant Signal Behav 5:469–472

Zhong R, Peña MJ, Zhou GK, Nairn CJ, Wood-Jones A, Richardson EA, Morrison WH, Darvill AG, York WS, Ye ZH (2005) Arabidopsis *Fragile Fiber8*, which encodes a putative glucuronyltransferase, is essential for normal secondary wall synthesis. Plant Cell 17:3390–3408

Zhong R, Demura T, Ye ZH (2006) SND1, a NAC domain transcription factor, is a key regulator of secondary wall synthesis in fibers of Arabidopsis. Plant Cell 18:3158–3170

Zhong R, Richardson EA, Ye ZH (2007a) The MYB46 transcription factor is a direct target of SND1 and regulates secondary wall biosynthesis in Arabidopsis. Plant Cell 19:2776–2792

Zhong R, Richardson EA, Ye ZH (2007b) Two NAC domain transcription factors, SND1 and NST1, function redundantly in regulation of secondary wall synthesis in fibers of Arabidopsis. Planta 225:1603–1611

Zhong R, Lee C, Zhou J, McCarthy RL, Ye ZH (2008) A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in Arabidopsis. Plant Cell 20:2763–2782

Zhong R, Lee C, Ye ZH (2010) Functional characterization of poplar wood-associated NAC domain transcription factors. Plant Physiol 152:1044–1055

Zhou CB, Chen AF (1985) Observation on the bionomics and hibernation of *Sesamia inferens* (Walker) in the northern part of Hainan Island. Insect Knowledge 22:199–201

Zhou GK, Zhong R, Richardson EA, Morrison WH, Nairn CJ, Wood-Jones A, Ye Z-H (2006) The poplar glycosyltransferase GT47C is functionally conserved with Arabidopsis Fragile Fiber8. Plant Cell Physiol 47:1229–1240

Zhou GK, Zhong R, Richardson EA, Himmelsbach DS, McPhail BT, Ye Z-H (2007) Molecular characterization of PoGT8D and PoGT43B, two secondary wall-associated glycosyltransferases in poplar. Plant Cell Physiol 48:689–699

Zhou J, Lee C, Zhong R, Ye ZH (2009) MYB58 and MYB63 are transcriptional activators of the lignin biosynthetic pathway during secondary cell wall formation in *Arabidopsis*. Plant Cell 21:248–266

Zhu XG, Long SP, Ort DR (2010) Improving photosynthetic efficiency for greater yield. Annu Rev Plant Biol 61:235–261

Ziegelhoffer T, Will J, Austin-Phillips S (1999) Expression of bacterial cellulase genes in transgenic alfalfa (*Medicago sativa* L.), potato (*Solanum tuberosum* L.) and tobacco (*Nicotiana tabacum* L.). Mol Breed 5:309–318

Zili Y, Puhua Z, Chengcai C, Xiang L, Wenzhong T, Li W, Shouyun C, Zuoshun T (2004) Establishment of genetic transformation system for *Miscanthus sacchariflorus* and obtaining of its transgenic plants. High Technol Lett 10:27–31

Zuo KJ, Qin J, Zhao JY, Ling H, Zhang LD, Cao YF, Tang KX (2007) Overexpression GbERF2 transcription factor in tobacco enhances brown spots disease resistance by activating expression of downstream genes. Gene 391:80–90

Chapter 13
Saccharinae Bioinformatics Resources

Alan R. Gingle and F. Alex Feltus

Abstract The primary goal of this chapter is to provide practical information for utilizing the array of *Saccharinae* bioinformatics resources that are presently available. The chapter begins with the description of a survey of *Saccharinae* bioinformatics resources that was undertaken early in 2010. Resources are categorized by life science area(s), available data types, and modes of data access. Navigating resources and searching for *Saccharinae* data is then described through a broad collection of search examples that cover categories ranging from maps, markers, and genomic sequence through transcriptome-, proteome-, and biochemistry-related data. Data integration, as means for providing answers to more complex biological questions, is discussed in terms of existing applications of reference genome sequence and possible future applications of co-expression network data.

Keywords Bioinformatics • Co-expression network • Database • Genome • Proteome • Saccharinae • Sorghinae • Transcriptome

1 Overview

While some have considered the *Saccharinae* to be composed of members from 13 genera, including *Saccharum*, *Miscanthus,* and *Imperata* (Clayton and Renvoize 1986), there is growing evidence that supports the merging of the *Saccharinae* and

A.R. Gingle (✉)
Plant Genome Mapping Laboratory, University of Georgia,
111 Riverbend Rd, CAGT, Rm 113, Athens, GA 30602-6810, USA
e-mail: agingle@uga.edu

F.A. Feltus
Department of Genetics and Biochemistry, Clemson University, Clemson, SC, USA
e-mail: ffeltus@clemson.edu

A.H. Paterson (ed.), *Genomics of the Saccharinae*, Plant Genetics and Genomics: Crops and Models 11, DOI 10.1007/978-1-4419-5947-8_13,

Sorghinae into a single clade (E. Kellogg, Chap. 1 of this volume). With this and inclusiveness in mind, the bioinformatic resource survey and descriptive text of this chapter include resources providing data for either *Saccharinae* or *Sorghinae*. Also, for the sake of conciseness, the *Saccharinae* and *Sorghinae* clades as well as bioinformatics resources, that provide related data, are referred to as "*Saccharinae*" in this chapter.

The spectrum of *Saccharinae* bioinformatics resources covers a broad range of life science areas and available data types of both genomic and phenotypic nature. For example, genetic map and diversity data are provided by genetics- and evolutionary biology-related resources, *Sorghum* genomic sequence views are available at the Joint Genome Initiative (JGI) Web resource and phenotypic data are provided by resources with focus areas, ranging from crop science to biochemistry. In addition, multi-organism resources, such as those at the National Center for Biotechnology Information (NCBI), provide a substantial amount of *Saccharinae* data that can be searched directly and are also available via links from other resources.

Searching bioinformatics resources often involves multiple data types from multiple resources. This renders manual searches, involving cutting and pasting results from one resource's interface into that of another, overly cumbersome. Many *Saccharinae* resources provide navigational aids that automatically link to and provide needed data for other resources, relieving users of this tedious task. Such links provide a useful degree of data integration and search automation; however, more can be done to expand the range of questions that users can ask and the ease with which users can obtain integrated views of the resulting data. For example, a bioinformatics resource can make its functionality available over the Web to facilitate discovery and use by client tools that automate complex searches and data processing into workflows. The incorporation of new Web technologies is an important component for enhancing the scope of possible searches and has been introduced in a similar chapter, focused on Gossypium bioinformatics resources (Gingle 2009).

Of course, the successful application of Web and information technology is dependent on the biological data to, at least, suggest interrelationships amongst the various data types, ranging from genomic through proteomic and beyond. The availability of whole genome sequence for *S. bicolor* (Paterson et al. 2009) is providing a reference structure for integrating the existing comparative and functional data. This potential is beginning to be realized at resources like JGI, NCBI and Phytozome (see Table 13.9 for URLs). These and other reference genome-based resources are included in the descriptions and workflows of Sect. 3. Co-expression networks, based on correlated gene expression levels, offer additional tools for identifying biologically meaningful relationships amongst genes and their regulatory mechanisms. The advent of next generation sequencing (NGS) technologies and related tools like RNA-Seq are likely to substantially increase the volume of the underlying gene expression data (Varshney et al. 2009; Wang et al. 2009) and their potential applications to data integration are described in Sect. 4.

For typical searches, life science users have formulated questions and are approaching the Web for answers. As part of this process, users need to select entry resources from amongst the battery of those available. Of course, knowledge of a

resource's available data types is essential for making this decision and, in addition, available data access methods may enter into the selection process. Accordingly, a survey of *Saccharinae* bioinformatics resources was undertaken by the authors and is the subject of the following section. This survey categorizes resources in terms of their life science focus areas, available data types and data access methods. A subsequent section employs these categories as a framework to illustrate search strategies as a function of data type. The recommendations contained in this chapter are, of course, based on the *Saccharinae* bioinformatics resources of the present. However, many of these resources are under active development and will likely expand their ranges of available data and, in some cases, data types as well.

2 The Spectrum of *Saccharinae* Informatics Resources

The primary goal of this chapter is to provide practical information for utilizing the array of *Saccharinae* bioinformatics resources that are presently available via the Web. To establish the setting, a survey of *Saccharinae* bioinformatics resources was undertaken by the author in early 2010. Given the taxonomic developments noted at the beginning of the previous section, the survey includes resources that are either *Saccharinae* or *Sorghinae* centric, focused on member genera or species as well as resources that provide data for either of these clades within a larger context. In the survey, resources were categorized by life science area and checked for available data types and access methods (e.g., data views, query tools, and Web services) as illustrated in Figs. 13.2 through 13.4.

2.1 Available Saccharinae Data by Genus

As noted above, the *Saccharinae* clade includes a number of genera and it is quite natural to inquire about the relative volume of available data by genus. The NCBI, by far the most comprehensive resource for Saccharinae data, was employed to gauge this by recording the total number of genome-, transcriptome-, and proteome-related records returned by its Entrez search tool for each member genus. This data, plotted in Fig. 13.1, illustrates the predominant interest in and data accumulation for Sorghum and then Saccharum, with Miscanthus a distant third, having returned records around the 0.1% level. Searches for other related genera returned near or below 100 records.

2.2 Life Science Area and Available Data Types

Saccharinae bioinformatics resources provide data and information from the perspectives of a wide range of life science areas. For example, the author's survey revealed resources with life science foci that covered areas as diverse as crop science

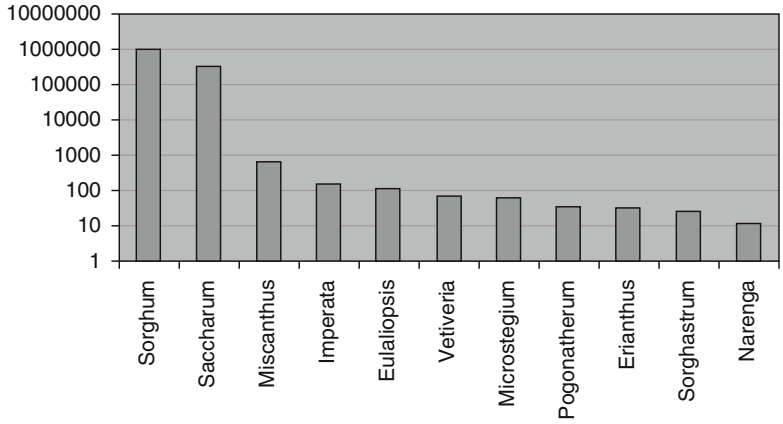

Fig. 13.1 Total number of genome-, transcriptome-, and proteome-related records returned by the National Center for Biotechnology Information's (NCBI) Entrez search tool for each of the Saccharinae genera. Record number is plotted logarithmically and only genera returning more than ten records are shown

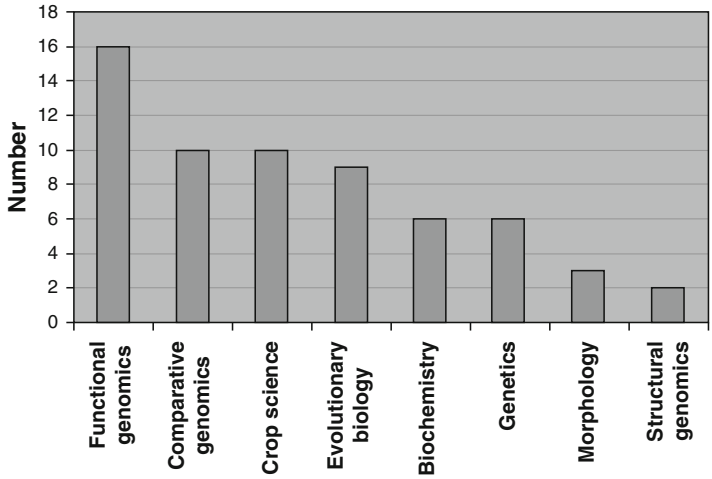

Fig. 13.2 Saccharinae bioinformatics resources classified by life science area

and structural genomics (Fig. 13.2). The survey included the full spectrum of *Saccharinae* bioinformatics resources ranging from those maintained by individual research labs or centers to large multi-organism resources with *Saccharinae* data such as those at NCBI and JGI.

Functional genomics is represented by the maximum number of resources followed by Comparative Genomics, the Crop Sciences and Evolutionary Biology.

Table 13.1 Available data types associated with *Saccharinae* bioinformatics resources

Category	Data type
Maps, markers, and genomic sequence related	Genetic map (individual loci and QTLs)
	Genetic marker
	Genomic sequence
	Physical map and BAC related
Diversity and evolution related	Genetic diversity
	Molecular evolutionary
	Taxonomic
Transcriptome related	Expression/microarray
	Expression/virtual
	Transcriptome/EST
Proteome related	Expression/protein
	Protein domains, sequence or structure
Phenotype related	Biochemical pathways
	Phenotypic traits
	Plant images
Agriculture related	Habitat and cultivation info
	Pedigree
	Performance trial
General information	Bibliographic information
General bioinformatics tool related	Sequence alignments

The availability of whole genome sequence and related resources for *S. bicolor* (Paterson et al. 2009) has substantially increased the volume and availability of genomic data. With respect to the Crop Sciences, the majority of the resources are devoted to sorghum variety trial data. Other areas, ranging from Biochemistry to Structural Genomics (3D protein structure determination) are represented as well.

Each of these life science areas has its own range of data types with their associated access, search, and presentation features. The data types that were found to be available in the recent survey of bioinformatics resources are listed in Table 13.1. They are grouped by categories that will be helpful to the example searches, appearing later in this chapter and a plot of bioinformatics resources grouped by these data types is shown in Fig. 13.3.

Each data type has its own range of modalities for data display. For example, genetic map data is typically displayed in standard genetic linkage map format as exemplified by Web resources like the Comparative Saccharinae Genomics Resource (CSGR), Gramene, and the Plant Genome Mapping Laboratory (PGML). Another example, whole genome sequence, can provide a reference for integrating loci, transcript, and other feature-/annotation-related data. Genome Browsers like GBrowse (Stein et al. 2002) or Vista (Couronne et al. 2003; Shah et al. 2004) take advantage of this inherent, biology-based, data integration through feature/annotation display tracks that parallel the reference genome sequence. Of course, Web displays for these and other data types have been developed, ranging in interactivity from static images to full-featured interfaces that respond to mouse click events. Unlike data

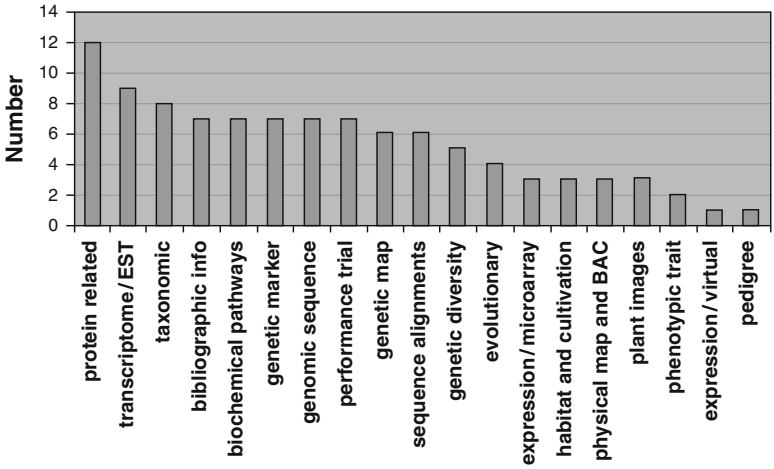

Fig. 13.3 *Saccharinae* bioinformatics resources classified by primary data types. Each resource providing a particular data type is counted even when multiple resources provide access to duplicative data sets

displays, search tools typically exhibit greater similarity across data types based on their common modes of entering search criteria (e.g., radio buttons as well as text, selection and check boxes). However, even they can exhibit data type distinctions especially when embedded in data display interfaces for convenience. In summary, a resource's available data types affect the kinds of data that can be queried, viewed, and downloaded and, more importantly, they also impact on the range of search criteria that are available for asking questions.

2.3 Methods of Data Storage and Accessibility

The manner in which data are stored and made accessible has a major impact on a resource's versatility with respect to searches and potential uses of its data by individual users and other resources. For example, data that are stored and made available as a PDF or word processor document are only searchable via simple text searches. Record-based flat files, such as those associated with spreadsheets (e.g., CSV), have greater utility since they can easily be uploaded into a relational database such as Access or MySQL for subsequent querying. Data stored in a searchable database (e.g., a relational database or RDB) has even greater potential both in terms of the range of possible searches and ways in which it can be integrated with the data of other resources. An illustration of the impact of the modes of data storage and accessibility is included in a survey of *Gossypium* bioinformatics resources (Gingle 2009) and it is germane in this context as well.

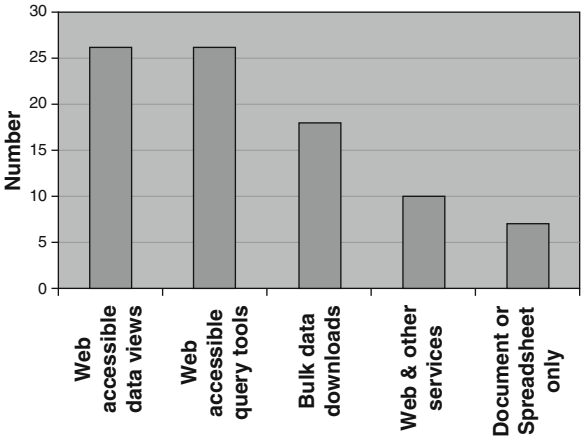

Fig. 13.4 *Saccharinae* bioinformatics resources classified by data access method

Saccharinae bioinformatics resources were also grouped by some common data access methods as part of the survey. The number of bioinformatics resources classified in this way is plotted (Fig. 13.4) for Web accessible data views, Web accessible query tools, PDF and other document availability, bulk data flat file availability, and Web and other services. The usefulness of Web services, software systems for resource-to-resource interaction and data exchange, are also covered in Gingle (2009). The impact of whole genomic sequence on genome, transcriptome, and proteome data integration has led to the commonality of Web accessible query tools to facilitate the increasingly more involved questions that biologists need answered. Bulk data downloads and Web services are also available in many instances. Resources providing PDF or other document downloads are especially common for crop sciences resources focused on Sorghum variety trial data.

2.4 Use of Ontologies

The genome ontology, GO (The Gene Ontology Consortium 2000, 2004), was found to be the most commonly used amongst the *Saccharinae* resources. This is a natural outcome of the availability of whole genome sequence for *S. bicolor* (Paterson et al. 2009) and the large volume of functional and comparative genomic data that are available for members of this and other clades from resources such as JGI, NCBI, and Phytozome (see Table 13.9 for URL). Other ontologies that are in limited use for *Saccharinae* data are related to plant structures and stages (PO; Jaiswal et al. 2005), traits (TO; Jaiswal et al. 2002), and molecular function (PANTHER; Thomas et al. 2003). Biochemical pathway ontologies such as those of the Kyoto Encyclopedia of Genes and Genomes (KEGG; Kanehisa et al. 2006) and BioCyc (Karp et al. 2005) as well as enzyme classifications are also in use.

The use of ontologies is likely to increase as additional genomes are sequenced and will benefit from the addition of phenotypic trait data that can be correlated with gene function and genetic diversity.

3 Searching for Saccharinae Data and Information

Optimal search methods and strategies for *Saccharinae* resources are dependent on the available data formats and search tools that are provided by the relevant resources. These depend on the primary data types that are involved with a particular search since many of the resource data formats, access methodologies, and search interfaces have developed around data type-specific requirements. In this section, example searches highlight these idiosyncrasies for many of the data categories and types as listed in Table 13.1. They are illustrated through workflow diagrams (Figs. 13.5, 13.6, 13.7, 13.8, 13.9, and 13.10) that highlight commonalities where they exist. Some resource-specific information and methodology are also included, especially for the more integrative searches that cover a broader range of data types.

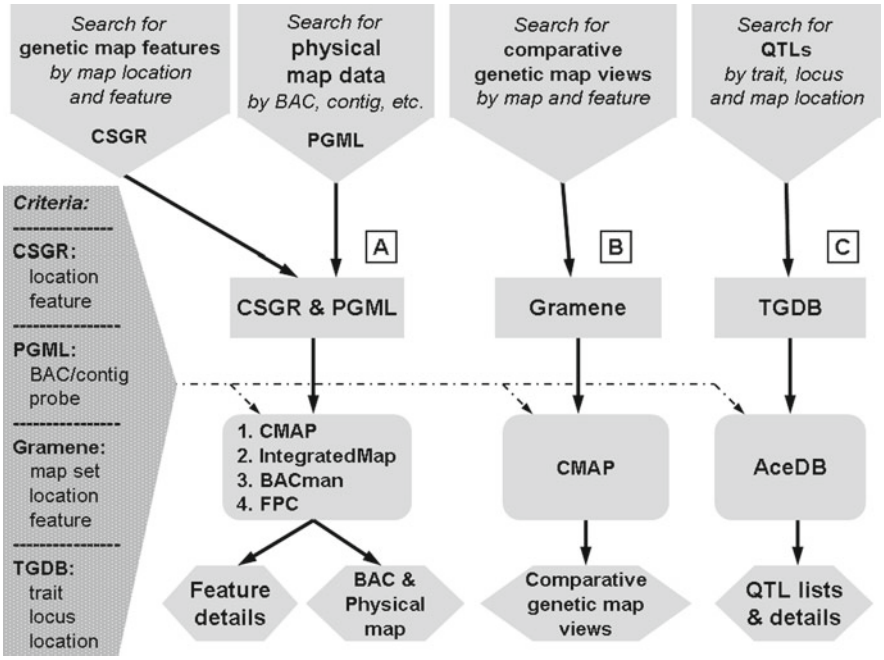

Fig. 13.5 Searches for genetic map features. Four resources that provide *Saccharinae* genetic map-related data were sampled by the author and are included. They are the Comparative Saccharinae Genomics Resource (CSGR) and the Plant Genome Mapping Lab (PGML) in *workflow A*, Gramene CMap in *workflow B*, and TropGENE-DB (TGDB) in *workflow C*. Some of the possible search criteria are indicated, by resource, on the *left*. Navigation paths are indicated by *solid arrows*, search criteria entry by *dashed lines* and data availability highlights in *hexagonal boxes*

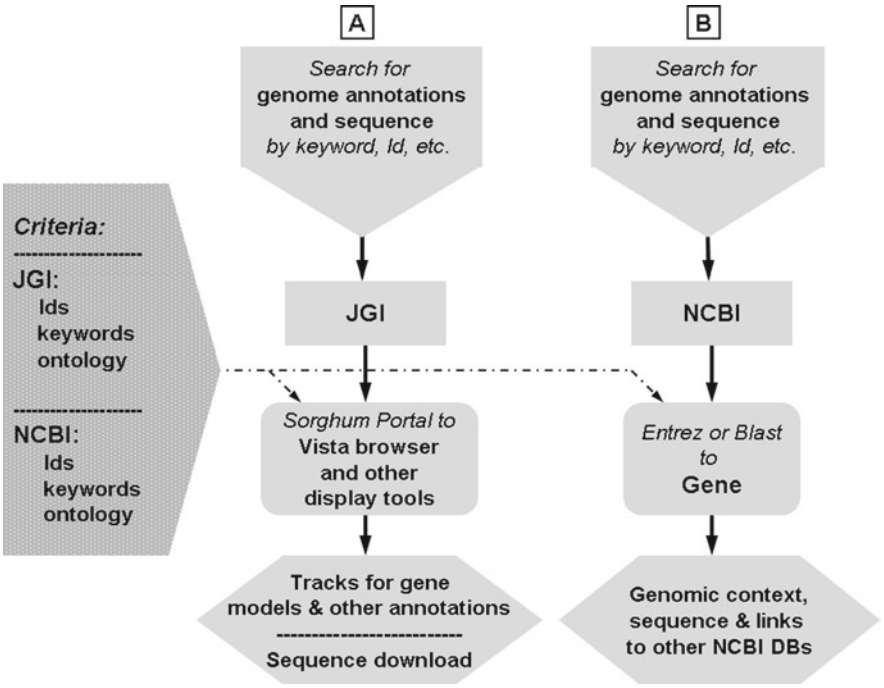

Fig. 13.6 Searches for genome annotations and genomic sequence. Two resources that provide related *Saccharinae* data were sampled by the author and are included. They are the Joint Genome Institute (JGI) Sorghum portal and the National Center for Biotechnology Information (NCBI) Gene database. At both JGI and NCBI, search criteria include feature Ids, annotation, and other keywords and ontology terms. At JGI, genome views are provided by a Vista-like genome browser with tracks for scaffold sequence, gene models, and other genome annotations (*workflow A*). At NCBI, a results page contains links to other NCBI databases, providing access to their full range of related data (*workflow B*). Some of the possible search criteria are indicated, by resource, on the *left*. Navigation paths are indicated by *solid arrows*, search criteria entry by *dashed lines* and data availability highlights in *hexagonal boxes*

3.1 Maps, Markers, and Genomic Sequence

Only *Sorghum* and *Saccharum* have substantial map resources such as genetic, genomic, and physical maps, as of this writing. *Sorghum* ranks first with its available genetic (Bowers et al. 2003; Mace et al. 2009), physical (Bowers et al. 2005) cytogenetic (Islam-Faridi et al. 2002; Kim et al. 2005), and genomic (Paterson et al. 2009) map resources. Genetic map and chromosomal genome sequence resources have been developed for *Saccharum* (Ming et al. 2002; Asano et al. 2004). So, the illustrated search workflows will focus on these genera. Physical map-related searches that involve BAC contig structure and hybridizations are also included in the illustrated workflows. We note that several additional resources are in progress, particularly for *Miscanthus* species.

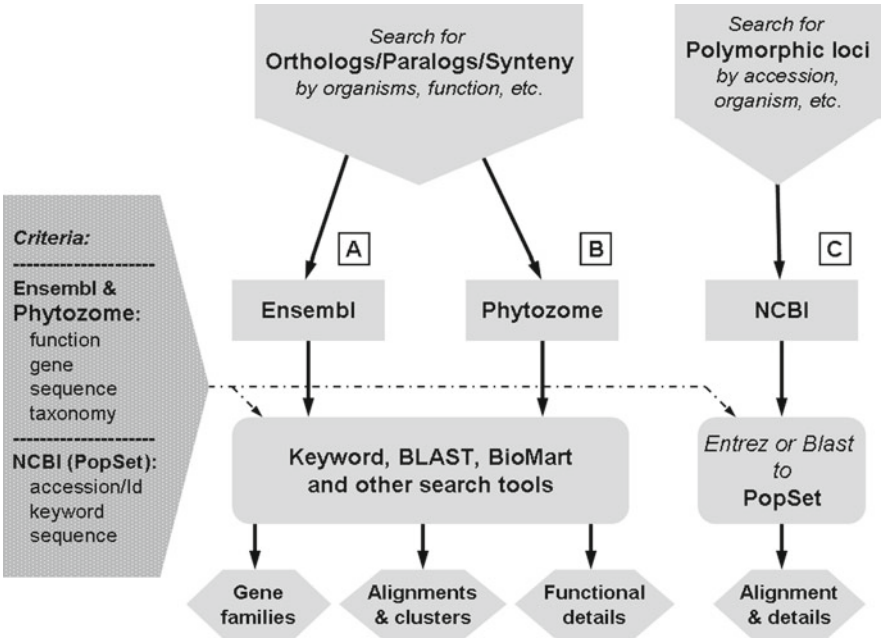

Fig. 13.7 Genetic diversity-related searches. Three resources that provide related *Saccharinae* data were sampled by the author and are included. They are Ensembl Plants, Phytozome, and the National Center for Biotechnology Information (NCBI). At Ensembl Plants and Phytozome, searches for gene family-related data are facilitated by keyword search tools, BLAST and other search tools such as BioMart (*workflows A* and *B*). At NCBI, searches for polymorphic loci are routed to their PopSet database display tools (*workflow C*). Some of the possible search criteria are indicated, by resource, on the *left*. Navigation paths are indicated by *solid arrows*, search criteria entry by *dashed lines* and data availability highlights in *hexagonal boxes*

Searches involving genetic or physical map features are illustrated in the diagrams of Fig. 13.5. Examples from four resources that provide *Saccharinae* genetic and / or physical map data are included. One resource, PGML, provides both genetic and physical map data for *Sorghum* while another, CSGR, provides search tools that access the PGML generated data for Sorghum. The related search examples are illustrated in Fig. 13.5A. At PGML, map data can be searched via their implementations of CMap for genetic and BACman (Estill et al. 2003) plus WebFPC (Soderlund et al. 2000) for BAC and physical map related data. Search criteria include map feature and location. PGML also provides Sorghum QTL data in a searchable format (see Table 13.3 for URL). At CSGR, their IntegratedMap and IntegratedSearch tools facilitate searches for sorghum genetic and physical map related data by map location and feature (Yang et al. 2005). These tools integrate data resources at CSGR and PGML to enhance the range of possible searches.

Another resource, Gramene CMap, provides comparative genetic map views for *Sorghum bicolor* to a range of other plant species. This is part of the Gramene genomics

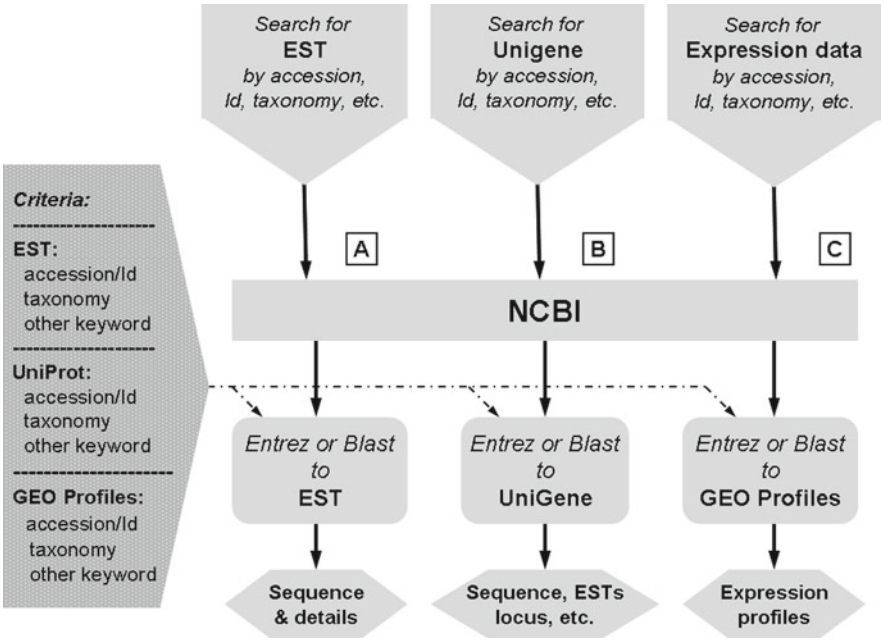

Fig. 13.8 Searches for EST, Unigene and expression data. Three NCBI databases that provide related *Saccharinae* data were sampled by the author and are included. They are the dbEST (EST), Unigene, and GEO/GEO Profiles. Searches for sequence, library, tissue, and other related information are facilitated by EST (*workflow A*). Searches for unigene sequence, member ESTs, and related locus info are facilitated by Unigene (*workflow B*). Searches for gene expression profiles and related platform and procedural info are facilitated by GEO/GEO Profiles (*workflow C*). All of the NCBI databases employ Entrez and Blast for keyword/text and sequence similarity-based searches, respectively. Some of the possible search criteria are indicated, by resource, on the *left*. Navigation paths are indicated by *solid arrows*, search criteria entry by *dashed lines*, and data availability highlights in *hexagonal boxes*

resource (Liang et al. 2008) that, while not *Saccharinae* centric, provides access to a range of *Sorghum* data, including intraspecific (Mace et al. 2009) and interspecific (Bowers et al. 2005) and QTL data (Feltus et al. 2006) for *S. bicolor*. Their implementation of CMap facilitates searches by criteria that include species, map set, map type, and feature (Fig 13.5B). In addition to genetic map views, their matrix interface allows users to search and display comparative map views of map feature pairs that are associated by a range of user configurable correspondence criteria.

Finally, TropGENE-DB (TGDB) provides genetic map, QTL, and other data types for sugarcane. Its QTL data for brix and stalk-related traits can be searched by trait name, linkage group, and correlation threshold (Fig. 13.5C). Their Web search tool contains selector boxes for the search criteria and returns associated QTLs with related data (Ruiz et al. 2004).

Searches for structural and functional genome annotations necessarily involve the genomic context whether displayed as static Web page views or user configurable

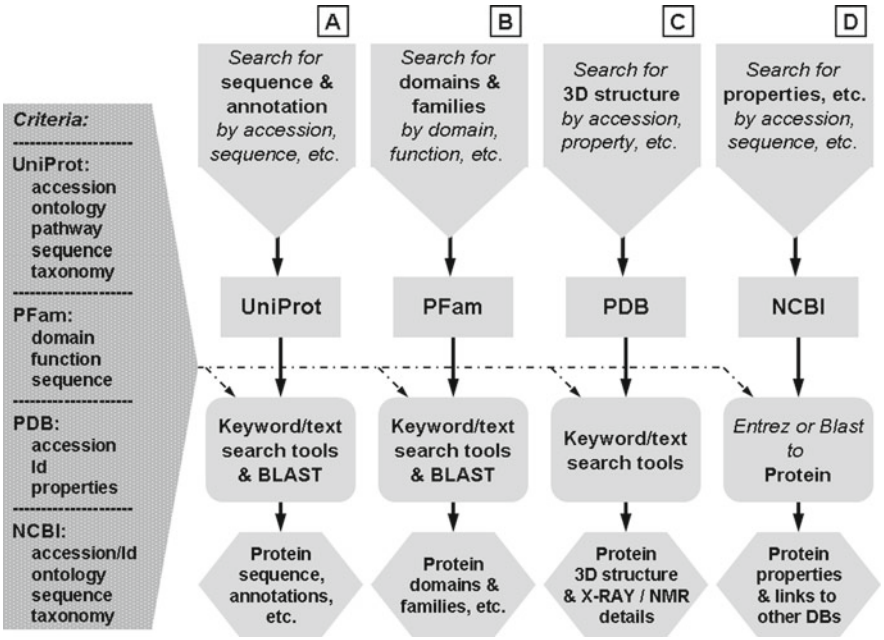

Fig. 13.9 Searches for proteome related data. Four resources that provide related *Saccharinae* data were sampled by the author and are included. They are UniProt, PFam, PDB, and the NCBI Protein database. Searches for protein sequence and related annotations at UniProt are illustrated in *workflow A*. Searches for protein families and related domains at PFam are illustrated in *workflow B*. Searches for 3D protein structures and related procedural information at the PDB are illustrated in *workflow C*. Searches for protein properties and links to other databases at NCBI-Protein are illustrated in *workflow D*. Some of the possible search criteria are indicated, by resource, on the *left*. Navigation paths are indicated by *solid arrows*, search criteria entry by *dashed lines*, and data availability highlights in *hexagonal boxes*

genome browser tracks. To date, the majority of *Saccharinae* genomic data has been generated for *Sorghum*. For example, an NCBI Entrez search with keyword, *Sorghum*, returned over 58 times as many genome/genomic sequence-related records as were returned by second-ranked keyword, *Saccharum*. Also, with the genome sequencing of *Sorghum bicolor* (Paterson et al. 2009) its genomic resources are the most advanced. Accordingly, searches for Sorghum genome annotation and sequence data are illustrated in the workflows of Fig. 13.6.

The JGI Sorghum portal enables searches for structural, functional, and comparative genome annotations by a versatile set of search criteria. Its search criteria range from key word searches with Boolean operations to those involving feature Ids and ontology terms. Both direct search criteria, related to genome annotations, and indirect criteria, related to homologous hit characteristics, are possible. Genome views are provided by a Vista like genome browser (Couronne et al. 2003; Shah et al. 2004) that displays tracks for scaffold sequence, gene models and other

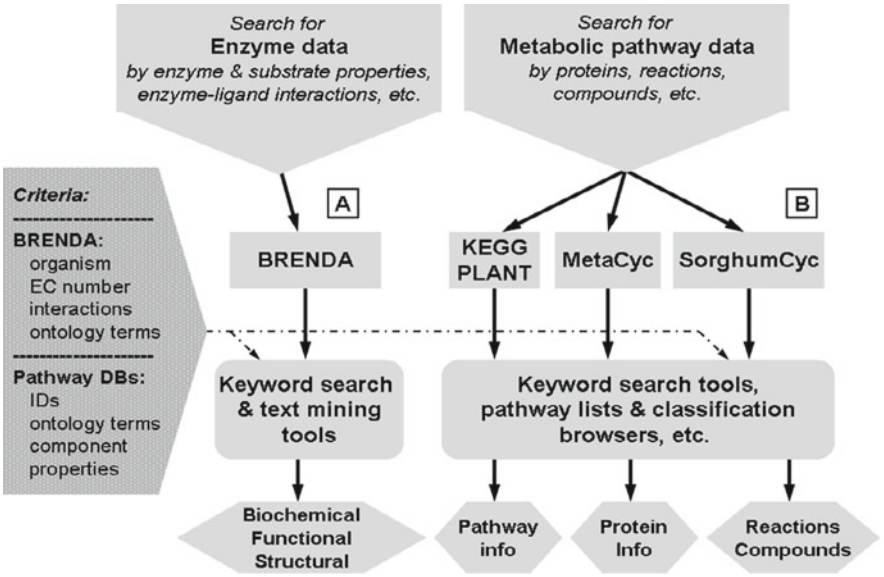

Fig. 13.10 Searches for biochemical enzyme and pathway related data. Four resources that provide Saccharinae data were sampled by the author and are included. They are the Braunschweig Enzyme Database (BRENDA), the Kyoto Encyclopedia of Genes and Genomes plant resource (KEGG PLANT), MetaCyc and SorghumCyc. At BRENDA, searches for enzyme and substrate properties as well as enzyme–ligand interactions are facilitated by keyword search and text mining tools (*workflow A*). At KEGG PLANT and the BioCyc-based resources, MetaCyc and SorghumCyc, searches related to biochemical pathways, reactions and compounds are facilitated by keyword search tools, pathway lists, and classification browsers (*workflow B*). Some of the possible search criteria are indicated, by resource, on the *left*. Navigation paths are indicated by *solid arrows*, search criteria entry by *dashed lines*, and data availability highlights in *hexagonal boxes*

genome annotations. Synteny plots for other taxa are also included. Menu options provide online BLAST, sequence data downloads and GO-, KEGG-, and KOG- (Tatusov et al. 2003) related information. A sample search workflow is illustrated in Fig. 13.6A. Another resource providing Sorghum genomic data, Phytozome, employs the GBrowse genomic browser (Stein et al. 2002) for genome views. This resource, focused on comparative genomics, is included in the example search workflows of Sect 3.2.

The NCBI gene database, a searchable database for genes from RefSeq (Wheeler et al. 2008), takes a somewhat different approach to the display and integration of genome annotation data. It contains a versatile search tool and a results page with graphical views of the genomic region, gene model, and other annotation data. The range of search criteria, from key word searches with Boolean operations to those involving feature Ids and ontology terms, is quite versatile. The results page contains links to other NCBI databases, providing access to the full range of related data at NCBI. For example, links to RefSeq and UniSTS allow users to obtain the

genomic sequence and any related STSs. This resource, however, does not employ a multitrack genome browser like GBrowse or Vista. A sample search workflow is illustrated in Fig. 13.6B.

3.2 Diversity and Evolution

Genetic diversity and evolutionary data for the *Saccharinae* are, for the most part, restricted to *Sorghum* and *Saccharum* (see Fig. 13.1 for NCBI records by genus). Collectively, resources for this data category focus on the evolutionary history of gene families, identification of syntenic blocks, evolutionary relatedness of populations, and taxonomy. The more involved searches relate to gene families, synteny, and population relatedness make up the example workflows of Fig. 13.7.

The Ensembl Genome project provides a wide range of genomic data, including comparative data (Flicek et al. 2008). The Ensembl "Plants" resource includes numerous entries for *Sorghum* and provides user customizable sequence alignment, ortholog and paralog identification, and gene tree views that indicate probable duplication and speciation events. Sequence alignments are editable via their implementation of Jalview (Waterhouse et al. 2009). In addition, they provide views of syntenic blocks. Their comparative resource facilitates search criteria that include organism, gene Id, and associated descriptive terms such as those related to biological function/process. Implementations of BLAST and BioMart (Durinck et al. 2005) enhance their search capabilities. An example search is illustrated in workflow A of Fig. 13.7.

The Phytozome resource facilitates comparative genomic studies amongst green plants, including *Sorghum*. Their Web tools facilitate searches of orthologous and paralogous gene clustering data to identify orthology/paralogy relationships and syntenic segments along with related duplication and speciation events. Related sequence alignment and genomic views are available as well as a range of annotation data. Search criteria include phylogenetic tree node, allowed/exclude/required organisms and gene family keywords such as Id or ontology term. Implementations of BLAST, BLAT (Kent 2002), and BioMart enhance their search capabilities. Like Ensembl, they employ Jalview for alignment editing and phylogenetic tree generation. An example search is illustrated in workflow B of Fig. 13.7.

The PopSet database at NCBI, focused on the evolutionary relatedness of populations, contains aligned sequences submitted as sets resulting from population, phylogenetic, or mutation studies (Wheeler et al. 2008). This database contains both nucleotide and protein sequence data, and, though *Sorghum* and *Saccharum* dominate the data records, ten *Saccharinae* genera are represented at the time of this writing. PopSet can be reached via their Entrez search tool and search criteria include organism name, Id/accession, and other key words. Their Entrez search tool is based on text strings that can optionally include Boolean operators (Benson et al. 2006). Record counts for each of their databases are displayed as links that ultimately lead to their data display tools. An example search is illustrated in workflow C of Fig. 13.7.

Two other resources for this data type are CSGR and Gramene and both provide *Sorghum* diversity data. At CSGR, their IntegratedMap and related search tools (Yang et al. 2005; Gingle et al. 2005) allow users to search for polymorphic loci-related diversity data associated with a study of Hamblin et al. (2004). Search criteria include genetic map location, locus, and polymorphism type. Their search tools link to and relay on NCBI PopSet for display of sequence alignment details. At Gramene Genetic Diversity, Web tools facilitate searching a broader range of *Sorghum* diversity data. Their tools provide related allele and germplasm data as lists and tables as well as in downloadable text format.

3.3 Transcriptome and Proteome

Saccharinae transcriptome and proteome assets are quite substantial. For example, an NCBI Entrez search, at the time of this writing, returned over 52 and 75,000 records in their dbEST and Protein databases, respectively. With respect to expressed sequence tag (EST) assembly/clustering, both *Saccharum* and *Sorghum* are represented in the Unigene sets at NCBI. However, their Genome Expression Omnibus (GEO) database contained a lesser volume of data; slightly over 3,000 gene expression profile records, exclusively for *Saccharum*. Other resources offer these data types for the *Saccharinae*, such as CSGR and the Dana-Farber Cancer Institute (DFCI) for EST/transcriptome related and the Protein Families database at the Sanger Institute (PFam), the Universal Protein Resource (UniProt) and the Protein Databank (PDB) for protein-related data. The resources for *Saccharum* and *Sorghum* dominate these data categories and, therefore, the example searches and illustrated workflows (Figs. 13.8 and 13.9) are focused on these genera.

Currently, NCBI's dbEST resource is the primary archive for *Saccharinae* transcriptome sequence resources and is included as one of the search examples illustrated in the workflows of Fig. 13.8A. This resource provides EST sequence and related information such as organism, tissue, treatment, etc. It, like other NCBI databases, employs Entrez for text/keyword searches and Blast for sequence similarity-based searches. While the vast majority of transcriptome sequence data resides at dbEST, NGS technologies are likely to make the NCBI Sequence Read Archive (SRA) for the short reads generated by these technologies an important source of transcriptome sequence data. At the time of this writing, only *Miscanthus* is represented with 366,448 reads (Swaminathan et al. 2010).

The natural progression of EST sequence data processing is toward sequence clustering and assembly to, amongst other goals, determine as much of the mRNA sequence as possible. As noted above, the NCBI resource, UniGene, provides unigene sets EST clusters and related sequence for two of the *Saccharinae* genera. Like dbEST, it employs Entrez for text/keyword searches and Blast for sequence similarity-based searches (Fig. 13.8B). As noted above, DFCI is another resource that provides this data type in the form of their gene indexes for *S. bicolor* and Sugarcane. Another possible goal/outcome of transcriptome sequence clustering is the gauging

of gene expression levels through the generation of digital expression profiles. CSGR provides digital gene expression data (Huang et al. 2005) for a limited number of *S. bicolor* EST libraries that were generated as part of a National Science Foundation plant genome project headed by A.H. Paterson (Pratt et al. 2005). It is expected that NGS technologies and the RNAseq approach to transcriptome profiling (Wang et al. 2009) will have a substantial impact on the future volume of expression data for the *Saccharinae* and in general.

At the time of this writing, a small volume of expression profile data is available for *Saccharum* at GEO and GEO Profiles. These data are based on four platforms, two sugarcane nylon arrays (GPL210 and GPL1329), one spotted array (GPL3799), and an Affymetrix Sugar Cane Genome array (GPL3844). The gene expression profiles relate to a range of studies of stress tolerance, brix content, sugar accumulation, etc. (Nogueira et al. 2003; Rocha et al. 2007; McCormick et al. 2008). The Affymetrix Sugar Cane Genome array-based datasets of McCormick et al. are also available at PLEXDB, a resource providing plant and plant pathogen gene expression data (Wise et al. 2007). Expression profiles (GEO Profiles), datasets and related expression array/platform information are available at GEO (Barrett et al. 2005; Edgar and Barrett 2006) and the related pages are interlinked so that a user can navigate from Entrez to either GEO or GEO Profiles after the initial Entrez text/keyword or Blast sequence similarity search (Fig. 13.8C).

Saccharinae proteome-related data resources range from the more general integrative Protein database at NCBI (Wheeler et al. 2008) to those that are more specialized like UniProt, PFam, and the PDB. Example searches for these resources are illustrated in the workflows of Fig. 13.9. UniProt is an archive for protein sequence and functional information annotated with gene ontology (GO) terms (The UniProt Consortium 2010; Schneider et al. 2005). At the time of this writing, it contains over 35,000 *Saccharinae*-related records, the vast majority for *Sorghum*. Possible search criteria include sequence similarity, ontology terms, and keywords related to pathway and ontology (Fig. 13.9A). Resources like PFam and InterPro are more focused on protein families and their related domains (Finn et al. 2010; Hunter et al. 2009). At the time of this writing, PFam contains over 600 *Saccharinae*-related records, primarily for *Sorghum* and *Saccharum* and with a small number for *Miscanthus*. Possible search criteria include sequence similarity and keywords related to protein domain and function (Fig. 13.9B).

The Protein Data Bank (PDB) is focused on providing protein 3D structures and related X-ray crystallographic (X-RAY) or nuclear magnetic resonance (NMR) structural determination parameters (Berman et al. 2007). At the time of this writing, six protein structures from *Sorghum* are available. Possible search criteria include keywords related to structural properties and X-RAY or NMR structural determination parameters (Fig. 13.9C). An example search for the Protein database at NCBI completes the workflow illustrations of Fig 13.9. As noted above, this resource is more general in scope and links extensively to other databases for more detailed information. It, like other NCBI databases, relies on Entrez for text/keyword searches.

3.4 Phenotype

Saccharinae resources provide a diverse range of phenotypic data, making biochemical-, diversity-, performance-, and stress-related data available in downloadable and/or searchable formats. While the largest portion of phenotype-related resources were found to be devoted to *Sorghum* performance trial data (see Fig. 13.3), all provide trial data in downloadable files for applications that have limited search capabilities (e.g., spreadsheets and PDF/Word documents). Other phenotypic data types are typically provided in more searchable formats. For example at TGDB, a phenotypic query tool facilitates searches for sugarcane susceptibility and tolerance to both abiotic and biotic stresses based on germplasm, as well as disease, pest, or abiotic stress type. At GRIN, searches for cultivation-, phenotype-, and general accession-related information by Id are facilitated by their search tool. Search criteria are entered by a combination of text entry and list selections and data is displayed in html/text format.

3.5 Biochemical

While biochemical data can be considered as phenotypic in nature, it is quite distinct from the range of qualitative and quantitative traits associated with crop performance and mutant morphotypes. Searches for biochemical compound-, reaction-, and pathway-related data are illustrated in the workflows of Fig. 13.10, which include example searches for enzyme-, reaction-, and biochemical pathway-related data. The Braunschweig Enzyme Database (BRENDA) provides a broad range of biochemical and molecular information on classified enzymes (Schomburg et al. 2004; Chang et al. 2009), including numerous entries for *Sorghum* and *Saccharum* as well as some for *Miscanthus*. The available information ranges from biochemical reaction through functional and organism to structural related data. Their automated text-mining-based databases, FRENDA and AMENDA, provide access to related literature in PubMed (Chang et al. 2009). Their software tools facilitate searches by any of their data fields that include organism, EC number, enzyme and substrate properties, enzyme–ligand interactions, and ontology terms (e.g., BTO and GO). Both simple, single criteria/data field, and combinatorial searches are possible and an example search is illustrated in workflow A of Fig. 13.10.

The Kyoto Encyclopedia of Genes and Genomes (KEGG) resource provides a range of biosynthetic pathway related data for a broad range of species, including plants (Kanehisa et al. 2006). At the time of this writing, a search by species yielded over 100 *Sorghum*-related records. Their plant-specific database, KEGG PLANT, provides information related to biosynthetic pathways, phytochemical compounds, and pharmacological applications of phytochemicals (Kanehisa et al. 2010). Their resource facilitates searches by a range of criteria that include organism, gene identifier, pathway alias enzyme number, reaction numbers, and pathway-related ontology. An example search is illustrated in workflow B of Fig. 13.10.

Both BioCyc-based resources, illustrated in Fig. 13.10B, employ the pathway tools developed by Karp et al. (2010) at SRI International and, therefore, their search and display functions are similar. MetaCyc is multi organism while SorghumCyc is, of course, a *Sorghum*-centric database of nonredundant metabolic pathways. Both allow pathway browsing and versatile search criteria that include compound, protein, reaction and pathway names, Ids, related ontology terms and properties of pathways, proteins, reactions or compounds. An example search is illustrated in workflow B of Fig. 13.10.

4 Integration of Multiple Data Types Using Gene Networks

In the last decade, there has been a rapid accumulation of genomic, genetic, and molecular knowledge for several plant species. This accumulation has begun for the *Saccharinae* and is predicted to continue robust growth for the foreseeable future. A primary driver for this prediction is the rapidly decreasing cost of each genomic data point, growing interest in the Saccharinae, and the promise that functional control points in *Sorghum*, *Miscanthus*, *Saccharum*, or other Saccharinae genomes can be quickly and reliably identified using the wealth of genomic data.

Genome-level data is naturally organized around a sequenced *reference genome* since the genome itself contains much of the information needed to describe the organism (although we are only just beginning to understand how to decipher it!). A static reference genome allows for the mapping of functional units such as DNA patterns that share homology to known genes. These mappings allow for inference of GO terms, protein domains, and pathway involvement. Direct measurements of allelic variation, the spectrum of nucleotide and/or structural differences among the members of a population, can easily be mapped to the reference genome. Dynamic changes in functional genomic information output including gene expression patterns measured with microarrays/RNA-seq or epigenetic modification to the primary DNA sequence can also be mapped to the reference genome although this becomes trickier in cases where there is overlapping information from multiple gene products encoded by a single locus. For the Saccharinae clade, the *Sorghum* genome provides a valuable reference for organization of data not only for itself, but also for the more complex and as-yet unsequenced genomes of *Miscanthus*, *Saccharum* and other species, akin to the way mammalian genomes can be organized around the human genome (UCSC genome browser). Eventually, of course, sugarcane, *Miscanthus*, and many other Saccharinae species will have their own reference genomes.

One hindrance to organizing genomic data around a static reference genome is that while positional interdependence between loci (i.e. linkage) is preserved, known and unknown relationships between loci are difficult to present in a useful manner. Most traits after all are controlled by multiple genomically dispersed loci, and even Mendelian gene products do not act in isolation within a cell. An exciting area of research is to converge seemingly disparate data types focused on individual

Fig. 13.11 A grass gene co-expression network. Shown is a fraction of a rice co-expression network which consists of ~4,000 rice genes interacting with ~50,000 relationships, built from 195 publically available Affymetrix Rice Genome Array experiments (GEO Acc#GPL2025). Each *dot* is a microarray probe set and each *line* is a statistically significant co-expression interaction. The network was drawn with Cytoscape software

loci and reference genome positions into easily accessible gene sets (modules) of cofunctional gene products that would be predicted to underlie agronomically and/or physiologically important traits. Such sets of interacting gene products can be easily modeled and displayed by using biological networks.

4.1 Network Biology

There are many possible interactions that can occur between genes. Examples include association by linkage (physical proximity), co-expression of gene products (mRNA, protein), and physical interaction (protein–protein, miRNA–target mRNA, *cis–trans*, etc.). Each of these interactions can be thought of as a graph in which binary interactions (genes A, B) are represented by nodes linked with a line (edge) than can be weighted and have directionality (Fig. 13.11). A collection of binary interactions (two nodes and one edge) can be discovered and represented as a gene network (Barabasi and Oltvai 2004). Once a network is constructed, it is possible to carve out highly connected gene sets (modules) that one may hypothesize to be cofunctional.

While biological networks are conceptually simple, being a collection of binary interactions with or without directionality, they can be extremely large and computationally challenging. Fortunately, there is growing collection of mature open-source software solutions (Suderman and Hallett 2007). Of note are Cytoscape (Cline et al. 2007), VisANT (Hu et al. 2008), and Bioconductor (Gentleman et al. 2004). Cytoscape and VisANT are excellent visualization tools that are extensible while Bioconductor is a collection of bioinformatics tools written in the statistical programming language, R, and perform many useful functions including network analysis and visualization.

4.2 *Co-expression Networks: An Example of a Biological Network*

In order to discover gene–gene dependencies with network analysis, one has to construct a network of dependencies. One of the more accessible data types that is available for several grasses are expression profiling experiments available from public repositories such as the GEO and Short Read Archive (SRA) at NCBI. Through a meta-analysis approach, it is possible to construct a network of statistically significant co-expression interactions between probe sets on an array. In plants, co-expression networks have been constructed for rice (Lee et al. 2009) and *Arabidopsis* (Loraine 2009; Wei et al. 2006; CressExpress). A portion of a rice co-expression network can be seen in Fig. 13.11. Co-expression networks can be subdivided into highly interacting gene modules using soft network approaches (Langfelder and Horvath 2008).

4.3 *Functionally Enriched Network Modules: Candidate Gene Sets Underlying Complex Traits*

Since genomic data is organized around fixed positions in a reference genome, it becomes possible to associate nodes in gene modules with functional units assigned to the genome position. For example, one can count the number of functional units (GO terms, protein domains, KEGG pathways, etc.) that are assigned to coordinates in the reference genome overlap with the gene products in the gene module. Using simple statistical techniques like Fisher's Exact test, it becomes possible to determine if the number of times a given functional unit is present in a set of genes (module) relative to the genomic background. There are numerous functional profiling tools (Al-Shahrour et al. 2007; da Huang et al. 2007; 2009a, b) that can achieve this task. Thus, one can compare a gene module identified in a sorghum co-expression network that is enriched in function X with a gene module identified in from a sugarcane that is also enriched in function X. Available mutant germplasm can be identified for those genes and the hypothesis can be tested in sorghum or sugarcane, whichever has the "best" genetic resources at the time, that the gene set underlies the complex trait.

5 *Saccharinae* Web Resource Survey: Table of Resources

This appendix contains tables of Web resources that provide *Saccharinae* data in each of the listed categories. A final table contains other URLs of projects and resources that are discussed in the text. Tables for each data category list resource name or abbreviation and URL. Resources providing multiple categories of data appear in multiple tables.

Table 13.2 Transcriptome related

Name and URL
Comparative Saccharinae Genomics Resource (CSGR) http://csgr.pgml.uga.edu
DFCI Gene Indexes http://compbio.dfci.harvard.edu/tgi/plant.html
NCBI (Entrez to EST, GEO, GEO Profiles, SRA and Unigene) http://www.ncbi.nlm.nih.gov/Entrez
Plant Expression Database (PLEXDB) http://www.plexdb.org
Plant Genome Database (PlantGDB) http://www.plantgdb.org/search/misc/PublicPlantSeq.php

Table 13.3 Maps, markers and genomic sequence

Name and URL
Comparative Saccharinae Genomics Resource (CSGR) http://csgr.pgml.uga.edu
Computational Biology and Functional Genomics Laboratory http://compbio.dfci.harvard.edu/tgi/plant.html
Ensembl Plants http://plants.ensembl.org
The European Molecular Biology Laboratory (EMBL Search) http://www.ebi.ac.uk/embl/
Gramene http://www.gramene.org/cmap
Joint Genome Institute Sorghum Portal http://genome.jgi-psf.org/Sorbi1/Sorbi1.home.html
NCBI (Entrez to Gene, Genome, Genome Project, GSS, Nucleotide, Probe, and UniSTS) http://www.ncbi.nlm.nih.gov/Entrez
Phytozome http://www.phytozome.net/sorghum
PlantGDB http://www.plantgdb.org/search/misc/PublicPlantSeq.php
Plant Genome Mapping Lab (BACman: *Sorghum* and *Saccharum*) http://www.plantgenome.uga.edu/bacman/BACManwww.php
Plant Genome Mapping Lab (Genetic maps and QTLs) http://www.plantgenome.uga.edu/cmap
Plant Genome Mapping Lab (Sorghum Physical map) http://www.plantgenome.uga.edu/fpc/WebAGCol/bicolor http://www.plantgenome.uga.edu/fpc/WebAGCol/propinquum
TropGENE-DB (Sugarcane) http://tropgenedb.cirad.fr

Table 13.4 Plant images

Name and URL
Germplasm Resources Information Network (GRIN) http://www.ars-grin.gov/npgs/searchgrin.html
Digital Flora of Texas Vascular Plant Image Library http://botany.csdl.tamu.edu/FLORA/gallery/gallery_query.htm
Plants Database http://plants.usda.gov

Table 13.5 Crop science related

Name and URL
Germplasm Resources Information Network (GRIN) http://www.ars-grin.gov/npgs/searchgrin.html
Plants Database http://plants.usda.gov
Missouri Sorghum variety trials http://varietytesting.missouri.edu/archive.htm
Nebraska Sorghum variety trials http://cropwatch.unl.edu/web/varietytest/sorghum
New Mexico Sorghum variety trials http://clovissc.nmsu.edu/variety-trials.html
Oklahoma Sorghum variety trials http://croptrials.okstate.edu/grain-sorghum
Purdue Crop Performance Program http://www3.ag.purdue.edu/agry/pcpp/Pages
Tennessee Sorghum variety trials http://varietytrials.tennessee.edu/sorghum.htm
Texas Sorghum variety trials http://varietytesting.tamu.edu/corn&grainsorghum/resources.htm

Table 13.6 Proteome related

Name and URL
Gramene http://www.gramene.org/protein
EMBL InterPro http://www.ebi.ac.uk/interpro
Joint Genome Institute Sorghum Portal http://genome.jgi-psf.org/Sorbi1/Sorbi1.home.html
NCBI (Entrez to 3D Domains and Protein) http://www.ncbi.nlm.nih.gov/Entrez
Pfam http://pfam.sanger.ac.uk
PlantGDB http://www.plantgdb.org/search/misc/PublicPlantSeq.php
Protein Databank (PDB) http://www.pdb.org
UniProt http://www.expasy.uniprot.org

Table 13.7 Diversity and evolution

Name and URL
Botanical Name Portal http://www.anbg.gov.au/ibis/speciesLinks.html
Comparative Saccharinae Genomics Resource (CSGR) http:///csgr.pgml.uga.edu
Ensembl Plants http://plants.ensembl.org
Genetic Resource Information Network (GRIN) http://www.ars-grin.gov/cgi-bin/npgs/html/index.pl
Gramene Genetic Diversity http://www.gramene.org/db/diversity/diversity_view
Institute of Genetic Diversity (IGD) http://www.sorghumdiversity.org/cgi-bin/sorghum/ searches/webform/moldiversity_search
Integrated Taxonomic Information System (ITIS) http://www.itis.gov
NCBI (Entrez to PopSet and Taxonomy) http://eutils.ncbi.nlm.nih.gov/Entrez
Phytozome http://www.phytozome.net/sorghum
Plant Genome Mapping Lab (synteny and comparative QTL data) http://www.plantgenome.uga.edu/cmap
Plants Database http://plants.usda.gov

Table 13.8 General information

Name and URL
Botanical Name Portal http://www.anbg.gov.au/ibis/speciesLinks.html
Comparative Saccharinae Genomics Resource (CSGR) http://csgr.pgml.uga.edu
Germplasm Resources Information Network (GRIN) http://www.ars-grin.gov/npgs/searchgrin.html
Gramene Sorghum http:www.gramene.org/species/sorghum/sorghum_intro.html
IPNI http://www.ipni.org/index.html
Plants Database http://plants.usda.gov

Table 13.9 Project and resource URLs mentioned in the text

Name and URL
Affymetrix: GeneChip Sugar Cane genome array page http://www.affymetrix.com/estore/browse/products.jsp?navMode=34000&productId=131508&navAction=jump&aId=productsNav
CMap http://gmod.sourceforge.net/cmap/index.shtml
CressExpress http://www.cressexpress.org
Cytoscape software http://www.cytoscape.org
Gene ontology (GO) http://www.geneontology.org/
Phytozome http://www.phytozome.net
Plant Ontology (PO) http://www.plantontology.org
R http://www.r-project.org
Sequence ontology (SO) http://www.sequenceontology.org
SRA http://www.ncbi.nlm.nih.gov/sra
Trait Ontology (TO) http://www.gramene.org/plant_ontology/ontology_browse.html#to
UCSC genome browser http://genome.ucsc.edu

Table 13.10 Biochemical and pathway related

Name and URL
Braunschweig Enzyme Database (BRENDA) http://www.brenda-enzymes.org
Encyclopedia of Metabolic Pathways (MetaCyc) http://metacyc.org
Kyoto Encyclopedia of Genes and Genomes (KEGG) http://www.genome.jp/kegg
NCBI (Entrez to PubChem databases) http://eutils.ncbi.nlm.nih.gov/Entrez
SorghumCyc http://ware.cshl.edu/sorghumcyc

Table 13.11 Phenotype related

Name and URL
Germplasm Resources Information Network (GRIN) http://www.ars-grin.gov/npgs/searchgrin.html
TropGENE-DB (SUGARCANE) http://tropgenedb.cirad.fr

References

Al-Shahrour F, Minguez P, Tarraga J, Medina I, Alloza E, Montaner D, Dopazo J (2007) FatiGO +: a functional profiling tool for genomic data. Integration of functional annotation, regulatory motifs and interaction data with microarray experiments. Nucleic Acids Res 35:W91–W96

Asano T, Tsudzuki T, Takahashi S, Shimada H, Kadowaki KI (2004) Complete nucleotide sequence of the sugarcane (*Saccharum officinarum*) chloroplast genome: a comparative analysis of four monocot chloroplast genomes. DNA Res 11(2):93–99

Barabasi AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. Nat Rev Genet 5(2):101–113

Barrett T, Suzek TO, Troup DB, Wilhite SE, Ngau W-C, Ledoux P, Rudnev D, Lash AE, Fujibuchi W, Edgar R (2005) NCBI GEO: mining millions of expression profiles—database and tools. Nucleic Acids Res 33:D562–D566

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL (2006) GenBank. Nucleic Acids Res 34:D16–D20

Berman H, Henrick K, Nakamura H, Markley JL (2007) The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. Nucleic Acids Res 35:D301–D303

Bowers JE, Arias MA, Asher R, Avise JA, Ball RT, Brewer GA, Buss RW, Chen AH, Edwards TM, Estill JC et al (2003) A high-density genetic recombination map of sequence-tagged sites for Sorghum, as a framework for comparative structural and evolutionary genomics of tropical grains and grasses. Genetics 165:367–386

Bowers JE, Arias MA, Asher R, Avise JA, Ball RT, Brewer GA, Buss RW, Chen AH, Edwards TM, Estill JC et al (2005) Comparative physical mapping links conservation of microsynteny to chromosome structure and recombination in grasses. Proc Natl Acad Sci USA 102:13206–13211

Chang A, Scheer M, Grote A, Schomburg I, Schomburg D (2009) BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009. Nucleic Acids Res 37:D588–D592

Clayton W, Renvoize S (1986) Genera *Graminum*. Her Majesty's Stationary Office, London

Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B et al (2007) Integration of biological networks and gene expression data using Cytoscape. Nat Protoc 2(10):2366–2382

Couronne O, Poliakov A, Bray N, Ishkhanov T, Ryaboy D, Rubin E, Pachter L, Dubchak I (2003) Strategies and tools for whole-genome alignments. Genome Res 13:73–80

da Huang W, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, Stephens R, Baseler MW, Lane HC, Lempicki RA (2007) The DAVID gene functional classification tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol 8(9):R183

da Huang W, Sherman BT, Lempicki RA (2009a) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37(1):1–13

da Huang W, Sherman BT, Lempicki RA (2009b) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4(1):44–57

Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, Huber W (2005) BioMart and bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics 21(16):3439–3440

Edgar R, Barrett T (2006) NCBI GEO standards and services for microarray data. Nat Biotechnol 24:1471–1472

Estill JC, Bowers JE, Marler BS, Paterson AH (2003) BACMAN: BAC data management for high throughput physical mapping of genomes. Plant Anim Microb Genome 11:300

Feltus FA, Hart GE, Schertz KF, Casa AM, Kresovich S, Abraham P, Klein PE, Brown PJ, Paterson AH (2006) Alignment of genetic maps and QTLs between inter- and intra-specific sorghum populations. Theor Appl Genet 112:1295–1305

Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunesekaran P, Ceric G, Forslund K, Holm L, Sonnhammer EL, Eddy SR, Bateman A (2010) The Pfam protein families database. Nucleic Acids Res 38:D211–D222

Flicek P, Aken BL, Beal K, Ballester B, Caccamo M, Chen Y, Clarke L, Coates C, Cunningham F, Cutts T et al (2008) Ensembl 2008. Nucleic Acid Res 36:D707–D714

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J et al (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5(10):R80

Gingle AR (2009) Gossypium bioinformatics resources. In: Paterson AH (ed) Genetics and genomics of cotton, plant genetics and genomics: crops and models 3. Springer Science, New York, pp 227–254

Gingle A, Huang Y, Wang H, Yang H (2005) New developments at CGGC: the integrated Web resource for Sorghum. Plant Anim Genome 8:276

Hamblin MT, Mitchell SE, White GM, Gallego J, Kukatla R, Wing RA, Paterson AH, Kresovich S (2004) Comparative population genetics of the panicoid grasses: sequence polymorphism, linkage disequilibrium and selection in a diverse sample of *Sorghum bicolor*. Genetics 167:471–483

Hu Z, Snitkin ES, DeLisi C (2008) VisANT: an integrative framework for networks in systems biology. Brief Bioinform 9(4):317–325

Huang Y, Pumphrey J, Gingle AR (2005) ESTminer: a Web interface for mining EST contig and cluster databases. Bioinformatics 21:669–670

Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L et al (2009) InterPro: the integrative protein signature database. Nucleic Acids Res 37:D224–D228

Islam-Faridi MN, Childs KL, Klein PE, Hodnett G, Menz MA, Klein RR, Rooney WL, Mullet JE, Stelly DM, Price HJ (2002) A molecular cytogenetics map of sorghum chromosome 1: fluorescence in situ hybridization analysis with mapped bacterial artificial chromosomes. Genetics 161:345–353

Jaiswal P, Ware D, Ni J, Chang K, Zhao W, Schmidt S, Pan X, Clark K, Teytelman L, Cartinhour S et al (2002) Gramene: development and Integration of trait and gene ontologies for rice. Comp Funct Genom 3(2):32–136

Jaiswal P, Ni J, Yap I, Ware D, Spooner S, Youens-Clark K, Ren L, Liang C, Zhao W, Ratnapu K et al (2005) Gramene: a bird's eye view of cereal genomes. Nucleic Acid Res 34:D717–D723

Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M (2006) From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res 34:D354–D357

Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Res 38:D355–D360

Karp PD, Ouzounis CA, Moore-Kochlacs C, Goldovsky L, Kaipa P, Ahren D, Tsoka S, Darzentas N, Kunin V, Lopez-Bigas N (2005) Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. Nucleic Acids Res 19:6083–6089

Karp PD, Paley SM, Krummenacker M, Latendresse M, Dale JM, Lee TJ, Kaipa P, Gilham F, Spaulding A, Popescu L et al (2010) Pathway tools version 13.0: integrated software for pathway/genome informatics and systems biology. Brief Bioinform 11(1):40–79

Kent WJ (2002) BLAT–the BLAST-like alignment tool. Genome Res 12(4):656–664

Kim JS, Islam-Faridi MN, Klein PE, Stelly DM, Price HJ, Klein RR, Mullet JE (2005) Comprehensive molecular cytogenetic analysis of sorghum genome architecture: distribution of euchromatin, heterochromatin, genes and recombination in comparison to rice. Genetics 171:1963–1976

Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9:559

Lee TH, Kim YK, Pham TT, Song SI, Kim JK, Kang KY, An G, Jung KH, Galbraith DW, Kim M et al (2009) RiceArrayNet: a database for correlating gene expression from transcriptome profiling, and its application to the analysis of coexpressed genes in rice. Plant Physiol 151(1):16–33

Liang C, Jaiswal P, Hebbard C, Avraham S, Buckler ES, Casstevens T, Hurwitz B, McCouch S, Ni J, Pujar A et al (2008) Gramene: a growing plant comparative genomics resource. Nucleic Acid Res 36:D947–D953

Loraine A (2009) Co-expression analysis of metabolic pathways in plants. Methods Mol Biol 553:247–264

Mace ES, Rami J, Bouchet S, Klein PE, Klein RR, Kilian A, Wenzl P, Zia L, Halloran K, Jordan D (2009) A consensus genetic map of sorghum that integrates multiple component maps and high-throughput diversity array technology (DArT) markers. BMC Plant Biol 9:13

McCormick A, Cramer M, Watt D (2008) Differential expression of genes in the leaves of sugarcane in response to sugar accumulation. Trop Plant Biol 1(2):142–158

Ming R, Liu SC, Bowers JE, Moore PH, Irvine JE, Paterson AH (2002) Construction of a Saccharum consensus genetic map from two interspecific crosses. Crop Sci 42:570–583

Nogueira FT, De Rosa VE, Jr MM, Ulian EC, Arruda P (2003) RNA expression profiles and data mining of sugarcane response to low temperature. Plant Physiol 4:1811–1824

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A et al (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556

Pratt LH, Liang C, Shah M, Sun F, Wang H, Reid S-P, Gingle AR, Paterson AH, Wing R, Dean R et al (2005) Sorghum expressed sequence tags identify signature genes for drought, pathogenesis, and skotomorphogenesis from a milestone set of 16,801 unique transcripts. Plant Physiol 139:869–884

Rocha FR, Papini-Terzi FS, Nishiyama MY Jr, Vêncio RZ, Vicentini R, Duarte RD, de Rosa VE, Jr VF, Barsalobres C, Medeiros AH et al (2007) Signal transduction-related responses to phytohormones and environmental challenges in sugarcane. BMC Genomics 13(8):71

Ruiz M, Rouard M, Raboin LM, Lartaud M, Lagoda P, Courtois B (2004) TropGENE-DB, a multitropical crop information system. Nucleic Acids Res 1(32):D364–D367

Schneider M, Bairoch A, Wu CH, Apweiler R (2005) Plant protein annotation in the UniProt knowledgebase. Plant Physiol 138:59–66

Schomburg I, Chang A, Ebeling C, Gremse M, Heldt C, Huhn G, Schomburg D (2004) BRENDA, the enzyme database: updates and major new developments. Nucleic Acids Res 32:D431–D433

Shah N, Couronne O, Pennacchio LA, Brudno M, Batzoglou S, Bethel EW, Rubin EM, Hamann B, Dubchak I (2004) Phylo-VISTA: interactive visualization of multiple DNA sequence alignments. Bioinformatics 20:636–643

Soderlund C, Humphray S, Dunham A, French L (2000) Contigs built with fingerprints, markers and FPC V4.7. Genome Res 10:1772–1787

Stein LD, Mungall C, Shu S, Caudy M, Mangone M, Day A, Nickerson E, Stajich JE, Harris TW, Arva A, Lewis S (2002) The generic genome browser: a building block for a model organism system database. Genome Res 12:1599–1610

Suderman M, Hallett M (2007) Tools for visually exploring biological networks. Bioinformatics 23(20):2651–2659

Swaminathan K, Alabady MS, Varala K, De Paoli E, Ho I, Rokhsar DS, Arumuganathan AK, Ming R, Green PJ, Meyers BC, Moose SP, Hudson ME (2010) Genomic and small RNA sequencing of *Miscanthus* × *giganteus* shows the utility of sorghum as a reference genome sequence for Andropogoneae grasses. Genome Biol 11(2):R12

Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN et al (2003) The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4:41

The Gene Ontology Consortium (2000) Gene ontology: tool for the unification of biology. Nat Genet 25:25–29

The Gene Ontology Consortium (2004) Gene ontology (GO) database and informatics resource. Nucleic Acids Res 32:D258–D261

The UniProt Consortium (2010) The universal protein resource (UniProt) in 2010. Nucleic Acids Res 38:D142–D148

Thomas PD, Kejariwal A, Campbell MJ, Mi H, Diemer K, Guo N, Ladunga I, Ulitsky-Lazareva B, Muruganujan A, Rabkin S et al (2003) PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. Nucleic Acids Res 31(1):334–341

Varshney RK, Nayak SN, May GD, Jackson SA (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends Biotechnol 27(9):522–530

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009) Jalview version 2 – a multiple sequence alignment editor and analysis workbench. Bioinformatics 25(9):1189–1191

Wei H, Persson S, Mehta T, Srinivasasainagendra V, Chen L, Page GP, Somerville C, Loraine A (2006) Transcriptional coordination of the metabolic network in *Arabidopsis thaliana*. Plant Physiol 142(2):762–774

Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Edgar R, Federhen S et al (2008) Database resources of the National Center for Biotechnology Information. Nucleic Acid Res 36:D13–D21

Wise RP, Caldo RA, Hong L, Shen L, Cannon EK, Dickerson JA (2007) BarleyBase/PLEXdb: a unified expression profiling database for plants and plant pathogens. In: Edwards D (ed) Plant bioinformatics – methods and protocols. Meth Mol Biol 406:347–363, Humana Press, Totowa, NJ

Yang H, Wang H, Gingle AR (2005) IntegratedMap: a Web interface for integrating genetic map data. Bioinformatics 21:2126–2127

Part III
Bridging Classical and Genomic Investigations of Sorghum Biology

Chapter 14
Bridging Classical and Molecular Genetics of Sorghum Plant Stature and Maturity

Patrick J. Brown and Andrew H. Paterson

Abstract Most initial sorghum introductions into the USA were too tall for mechanical grain harvest and flowered late or not at all at temperate latitudes. A small number of spontaneous mutations for dwarf stature and early maturity arose in farmers' fields and were quickly disseminated. Several subsequent sorghum introductions were determined by allelism tests to carry recessive alleles at the same loci. These early experiments led to the conclusion that there are four major loci for dwarfing in sorghum (*Dw1–Dw4*) and four major loci for maturity (*Ma1–Ma4*). The relatively simple inheritance of these important agronomic traits was later exploited by the Sorghum Conversion Program, which introgressed QTL for dwarfing and early maturity into hundreds of exotic lines through a strategy of backcrossing with phenotypic selection. To date, molecular work has identified just two of these classical dwarfing/maturity loci in sorghum: *Dw3* encodes an MDR-class auxin efflux carrier, and *Ma3* encodes a phytochrome B. In this chapter, we provide an overview of current understanding of the genetic architecture of plant height and flowering time in sorghum, including recent progress in mapping additional loci beyond the classical *Dw1–Dw4* and *Ma1–Ma4*. We also discuss prospects for identifying the remaining major loci for height and maturity in sorghum using both linkage- and association-based methods. Most of our understanding of sorghum phenotypic variation currently comes from studies in dwarf grain sorghum genetic backgrounds in

P.J. Brown (✉)
Department of Crop Sciences, 1408 Institute for Genomic Biology,
University of Illinois, 1206W. Gregory Avenue, Urbana, IL 61801, USA
e-mail: pjb34@illinois.edu

A.H. Paterson
Plant Genome Mapping Laboratory and Departments of Crop
and Soil Science, Plant Biology, and Genetics, University of Georgia, Athens, GA, USA

A.H. Paterson (ed.), *Genomics of the Saccharinae*, Plant Genetics and Genomics:
Crops and Models 11, DOI 10.1007/978-1-4419-5947-8_14,

which variation for other developmental traits may be masked. Growing interest in C4 grasses as bioenergy crops is likely to necessitate study of a broader range of sorghum genetic backgrounds.

Keywords Plant height • Flowering time • Photoperiod-sensitivity • Thermo-sensitivity • Introgression • Complimentary dominance • QTL • Association mapping

1 Classical Genetics

The timing of flowering is of central importance to drought avoidance, adapting many grasses to seasonal fluctuation in rainfall in their native environments. Most grasses originating in the semiarid tropics are photoperiod-sensitive and flower under short days (less than 12 h), coordinating seed development with favorable rainfall, temperature, and solar radiation conditions (Harper 1977). To adapt these and other cereals to temperate agriculture, photoperiod insensitive (day-neutral) mutations have been selected that cause plants to flower based on accumulation of "heat units" rather than under a specific daylength.

The timing of flowering of a plant is often closely related to both leaf/phytomer number and the overall height achieved by the plant's vegetative parts. In sorghum, a distinction is typically made between maturity genes, which influence the number of phytomers produced prior to floral initiation, and the dwarfing genes, which influence internode elongation but not the number of phytomers. Whereas dwarfing genes influence only plant height, maturity genes influence both plant height and flowering time. Most of our understanding of the genetic control of plant height and flowering time in sorghum derives from studies of dwarf grain sorghum grown at temperate latitudes. Specifically, this chapter draws heavily from work by Quinby and colleagues at the Texas Agricultural Experiment Station during the middle part of the twentieth century (Quinby 1974).

The first sorghum to be grown in the United States was probably broomcorn, which was cultivated by Benjamin Franklin as a curiosity (Maunder 1999). Sweet and forage sorghums were also introduced to the U.S. earlier than grain sorghums. However, mutations at height and flowering time loci were not reported until sorghum began to find use as a grain crop, perhaps because they were less useful for sweet and forage sorghums. One of the earliest varieties of sorghum to be grown for grain in the USA, and in many ways the variety that gave rise to sorghum genetics, was Milo. Soon after its introduction, it had spawned mutations at three maturity loci (*Ma1*, *Ma2*, and *Ma3*) and two dwarfing loci (*Dw1* and *Dw2*). This allowed the creation of the Milo maturity standards, which are near-isogenic lines for combinations of the four classical *Ma* loci in a three-dwarf (*dw1 dw2 Dw3 dw4*) background (Table 14.1). Crossing new sorghum introductions to the Milo maturity standards allowed their complement of *Ma* alleles to be deduced. Recessive alleles at the remaining *Ma* and *Dw* loci were found in other early introductions: Hegari already

Table 14.1 The Milo maturity standards

Line	Genotype
100M	Ma1Ma2Ma3Ma4
SM100	ma1Ma2Ma3Ma4
90M	Ma1Ma2ma3Ma4
SM90	ma1Ma2ma3Ma4
80M	Ma1ma2Ma3Ma4
SM80	ma1ma2Ma3Ma4
60M	Ma1ma2ma3Ma4
SM60	ma1ma2ma3Ma4
58M	Ma1Ma2ma3^{R}Ma4
44M	Ma1ma2ma3^{R}Ma4
38M	ma1ma2ma3^{R}Ma4

Table 14.2 Spontaneous mutant and native recessive alleles at the *Ma* and *Dw* loci

	Source of recessive allele(s)			
Locus	Spontaneous	Native	Gene	Notes
Ma1	Milo	Kafir	?	The major photoperiod-sensitivity locus
Ma2	Milo	Hegari	?	Complex interaction with *Ma1*
Ma3	Milo (*ma3*) Hegari (*ma3*) Milo (ma3R)	–	Phytochrome B	–
Ma4	–	Hegari	?	–
Dw1	Milo	–	?	–
Dw2	Milo	–	?	Linked to *Ma1*
Dw3	Kafir	–	MDR-class auxin efflux carrier	Unstable
Dw4	–	Most lines except broomcorn	?	–

carried a *ma4* allele when it was introduced from Africa, Broomcorn is the source of dominant *Dw4*, and the unstable *dw3* mutation occurred in a Kafir background.

The strength of the canonical *Ma1–4*/*Dw1–4* model lay in the fact that as additional maturity and dwarfing loci arose by mutation or were introduced in new varieties, these loci all proved to be allelic to existing loci. Mutations at the *Ma3* locus arose twice in Milo (once to the complete loss-of-function *ma3R* allele) and once in Hegari. Most Kafirs were already recessive at *ma1* when introduced from Africa, and likewise Hegari was already segregating for *ma2* when introduced (Quinby 1974). Mutant alleles at each of the eight classical *Ma* and *Dw* loci are summarized in Table 14.2.

The *Ma1–4*/*Dw1–4* model made simple, accurate predictions. Lateness at the maturity loci showed complete dominance. Mutation at the single *Ma1* locus was

sufficient to confer temperate-adaptedness. Tallness at the dwarfing loci showed incomplete dominance, with heterozygotes slightly shorter than tall-allele homozygotes. In the case of both maturity and dwarfing, the effect of any single mutation was usually greater in a late/tall background than in a background that already carried some recessive alleles. For example, the height difference between three-dwarf and four-dwarf near-isogenic lines was often less than 10 cm.

However, the *Ma1–4*/*Dw1–4* model also raised questions that have yet to be fully answered. Specifically, it was observed that *Ma1*/*ma1* heterozygotes flower later than either homozygote, but this only holds true only in populations where *ma2* is recessive (Quinby 1974). In populations with a *Ma2* allele, *Ma1*/*ma1* and *Ma1*/*Ma1* individuals are indistinguishable. Heterozygotes at *Ma2*, *Ma3*, and *Ma4* never show such overdominance in any genetic background. Equally puzzling is the observation that the effect direction of *Ma2* is dependent on *Ma1*: in a *Ma1* background, *Ma2* causes lateness, whereas in a *ma1* background, *Ma2* causes earliness.

2 The Sorghum Conversion Program

Mutations at the *Ma1–4* and *Dw1–4* loci, along with the discovery of cytoplasmic male sterility created by the interaction of kafir nuclear genes with a milo cytoplasm, provided breeders with an early toolkit with which to create early, dwarf, hybrid grain sorghum suitable for mechanical harvest. However, as grain sorghum became more popular and grew in acreage there was a growing apprehension that the limited number of initial sorghum introductions had given rise to a crop with a very narrow genetic base. Genetic diversity for resistance to biotic and abiotic stresses was particularly perceived to be in short supply. Although there was known to be tremendous worldwide genetic diversity in sorghum, the vast majority of this diversity was "locked up" in photoperiod-sensitive exotic varieties that wouldn't flower at temperate latitudes. Differences between flowering regimes are also a hindrance to using improved temperate germplasm in the tropics.

The solution to this problem was the Sorghum Conversion Program (SCP; Stephens et al. 1967), in which a diverse panel of exotic sorghum lines was introgressed with dwarfing and early maturity loci from a common donor (Fig. 14.1). All crosses were done by hand-emasculation in Puerto Rico, where photoperiod-sensitive and -insensitive lines would flower in synchrony. Since photoperiod-sensitivity in sorghum is usually dominant, F_1 generations also had to be grown and selfed in the tropics. F_2 and subsequent generations of progeny in which selection was to be performed had to be grown in a temperate location (Texas) in order to select short, early segregants. Crosses were made using the elite, temperate-adapted donor line as the female, with the exception of the final cross, which used the exotic parent as the female in order to restore each line to its native cytoplasm.

The SCP has converted more than 673 exotic sorghums to day-neutral flowering and combine height by backcross introgression of photoperiod-insensitive and dwarfing mutations (Rosenow and Clark 1987; Rosenow et al. 1996, 1997; Dahlberg

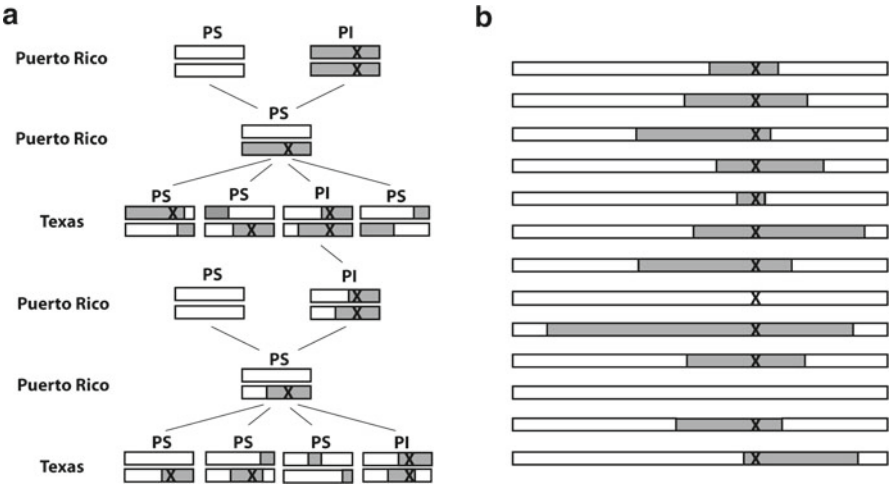

Fig. 14.1 The Sorghum Conversion Program. (**a**) Schematic showing the initial cross and first backcross during the creation of a sorghum converted (SC) line. Exotic and elite parental haplotypes are colored *white* and *gray*, respectively, and the "X" indicates a hypothetical recessive locus conferring photoperiod-insensitivity. PS and PI indicate photoperiod-sensitive and photoperiod-insensitive individuals. For simplicity, only a single chromosome is shown. (**b**) Hypothetical haplotype structure for 13 SC lines around a locus conferring dwarfism or photoperiod-insensitivity. Most SC lines are expected to carry the introgressed QTL, though the boundaries of the introgressed elite haplotype will vary between lines. Some lines may carry an independent mutation in the QTL locus (*Line 8*) or a wild-type allele (*Line 11*)

et al. 1998). The "converted" genotypes provide temperate-zone breeders with "user-friendly" access to exotic germplasm. Creation of Sorghum Conversion (SC) lines is ongoing, and typically involves five backcrosses, such that each SC line would be expected to be more than 98% identical with its exotic progenitor in the absence of selection. Depending on the number of selected loci in a given line and their linkage to each other, the actual percent identity between an SC line and its exotic progenitor may be much lower. The nonidentical portions of the genome are expected to comprise introgressed QTL for dwarfing and early maturity from the common donor, which is usually the four-dwarf Tx406.

It is worth noting that although the SCP was implemented with the *Ma1–4*/*Dw1–4* model in mind, the introgressed QTL in the SC lines are not necessarily limited to *Ma1–4* and *Dw1–4*. Any loci that confer tallness or photoperiod-sensitivity may be converted. Thus, evidence of conversion in SC lines can be used as evidence for maturity or dwarfing loci in the absence of any phenotypic data. If we assume that each SC line was backcrossed five times and retains just 2% Tx406 genotypes (1 in 50 markers) in unselected regions of the genome, then the chance that two SC lines will retain the same Tx406 marker in the absence of selection is just 1 in 2,500. If three or more SC lines are converted for a given marker, then this region is almost certainly being selected.

Molecular analyses of the SC lines have the potential to map QTL for plant height and flowering time, in the absence of phenotypic data, by identifying regions in which the elite donor genotype is being maintained through selection. Lin et al. (1995) pioneered this approach, finding conversion in the vicinity of the *Ma1*/*Dw2* loci on chromosome 6 for all nine lines studied. While the detection of other converted regions was limited by their very small sample size (just nine SC lines and their exotic progenitors), they did find evidence of conversion in nine additional unlinked genomic regions, five of which were closely associated with QTLs for either height or flowering.

Klein et al. (2008) focused specifically on chromosome 6, genotyping 87 markers spread evenly across the chromosome in a panel of 69 inbreds that included 18 SC lines. As expected, the milo-derived *Ma1*/*Dw2* haplotype of the elite donor line Tx406 was usually present in the converted lines. In some cases this donor haplotype extended across nearly the entire chromosome, which could reflect either suppression of recombination or selection of additional loci on chromosome 6. Some SC lines retained a distinct kafir haplotype around the *Ma1* locus yet displayed a photoperiod-insensitive (*ma1*) phenotype, giving credence to early observations that recessive mutations in the *Ma1* locus occurred independently in milo and kafir backgrounds (Table 14.2). Such genetic heterogeneity may complicate association-based approaches to gene identification.

3 Molecular Genetics

3.1 Cloned Loci

Just two of the eight classical maturity/dwarfing loci in sorghum have been cloned to date. The first locus to be cloned was *Ma3*, which encodes a phytochrome B (Childs et al. 1997). The authors used two milo maturity standards that are near-isogenic lines for *Ma3* (100M is *Ma3*; 58M carries the complete loss of function *ma3R* allele) to generate a population with simple Mendelian segregation for flowering time to and map markers that cosegregated with the putative *Ma3* locus. Previous work had established that *ma3R* genotypes contained only one of the two phytochrome proteins present in green leaves of both *Ma3* and *ma3* genotypes (Childs et al. 1992), and that the *ma3R* allele conferred early flowering irrespective of photoperiod treatment (Childs et al. 1995). Therefore, it was known that *ma3R* was a phytochrome-defective mutant, but not known whether the *Ma3* locus encoded a phytochrome directly or whether it encoded some other factor necessary for phytochrome protein accumulation. With this evidence in hand, the authors cloned cDNAs of phytochromes A, B, and C (*phyA–C*) and placed them on the same genetic map as the markers that cosegregated with *Ma3*. *PhyB* was observed to map in the same genetic location as the *Ma3* markers, and sequencing the *PhyB* gene showed that the *ma3R* mutant contained a frame-shift mutation.

The *Dw3* gene encodes an MDR-class auxin efflux carrier (Multani et al. 2003). Success in cloning this gene was due in part to the fact that it is unstable, producing wild-type revertants at low frequency, and in part to the fact that a mutation in the

orthologous maize gene phenocopies the sorghum mutation. The maize locus, *brachytic2* (*br2*), was cloned by transposon-tagging using the Mutator system. A *br2* probe was then used to isolate DNA from a homologous locus in sorghum, *Dw3*. Sequencing the *Dw3* gene in both *dw3* mutants and wild-type *Dw3* revertants revealed an 882 bp tandem duplication in the fifth exon that was present in all the mutants and none of the revertants. Instability at the *dw3* locus therefore appears to be due to reversion of the tandem duplication to a single copy through unequal crossing-over; further evidence for this was provided by the identification of *dw3* mutant plants that had accumulated a third copy of the 882 bp sequence (Multani et al. 2003). The authors also reported the discovery of an apparently stable *dw3* allele, which has significant commercial value since *dw3* instability is a cosmetic nuisance in grain sorghum production.

The *dw3*/*br2* loci are of agronomic interest for several reasons. First of all, their dwarfing effect is limited to the lower stalk internodes. Several studies have shown that the apical internodes, including the grain-bearing inflorescence, are actually longer in *dw3* mutant plants, and adequate exsertion of the panicle from the flag leaf is seldom a problem in *dw3* backgrounds (Schertz 1973; Brown et al. 2008). Second, stalk strength is increased in a *br2* background, which was demonstrated genetically by growing double-mutants for *br2* and *brittle stalk2* (*bk2*) to maturity under field conditions; stalks of *bk2* single mutants are easily snapped by gusts of wind or other slight disturbances (Sindhu et al. 2007). Increased stalk strength in *dw3* mutants is accompanied by extra layers of parenchyma cells in *dw3* internodes. Perturbing auxin transport clearly affects plant development, and in doing so may also influence biomass composition.

3.2 *Quantitative Trait Locus and Association Studies*

An early sorghum quantitative trait locus (QTL) study was performed by Pereira and Lee (1995), who measured plant height in a population of ~150 F_2s from a cross between CK60 (*dw1 Dw2 dw3 dw4*) and PI229828, a sudangrass line. They found three major QTL, one on chromosome 7 corresponding to *Dw3*, and two more on chromosomes 9 and 10, which may correspond to *Dw1* and *Dw4*. Tallness was dominant at the first two loci, but for the QTL on chromosome 10, the short allele was dominant. A fourth, smaller height QTL was found on chromosome 6, which could reflect either an allelic series at *Dw2* or a height QTL linked to the *Dw2* haplotype. The authors also found evidence for nonadditive interaction between the *Dw3* QTL and the height QTL on chromosome 9.

Lin et al. (1995) measured both plant height and flowering time in 370 F_2 individuals from an interspecific cross between *Sorghum bicolor* and *S. propinquum*, and found six QTL for plant height and three for flowering time. Three of the height QTL map to the approximate locations reported by Pereira and Lee (1995). The chromosome 10 height QTL was not detected, but additional QTL for plant height were detected on chromosomes 1 and 3. A flowering time QTL with an enormous effect ($a = 40$ days), explaining ~85% of the total phenotypic variation and delaying flowering of both heterozygous and homozygous genotypes until daylength dropped

below 12 h, was detected on chromosome 6, linked to a large-effect QTL for plant height. Since *Ma1* and *Dw2* were known from classical genetic work to be linked, this study established that *Ma1* and *Dw2* were on chromosome 6.

A population of ~140 (BTx623 × IS3620C) RILs has been used by several groups to map sorghum QTL for plant height and flowering time (Hart et al. 2001; Brown et al. 2006). The major QTL that were detected in both these studies include a plant height QTL linked to *Dw3* on chromosome 7 and a flowering time QTL on chromosome 9; the latter maps close to QTL previously detected for both flowering time (Lin et al. 1995) and plant height (Pereira and Lee 1995). Klein et al. (2001) used a different population of 125 (Sureno × RTx430) RILs to map QTLs for plant height, and again detected a *dw3*-linked QTL, as well as a novel plant height QTL on chromosome 4. Finally, Murray et al. (2008) used a population of 176 (BTx623 × Rio) RILs to map a *dw3*-linked QTL for plant height, a *Ma1*/*Dw2*-linked QTL for flowering time, and colocalizing QTL for both traits on chromosome 9, in approximately the same location as plant height and flowering time QTL previously reported (Pereira and Lee 1995; Lin et al. 1995; Hart et al. 2001; Brown et al. 2006). Clearly, the repeated discovery of a few QTL in corresponding locations supports the classical genetic interpretation that there are a limited number of major loci affecting plant height and flowering time in sorghum, or at least in grain sorghum.

Several recent studies have also mapped genomic regions involved in plant height and flowering time variation in sorghum using association methodology (Brown et al. 2008; Murray et al. 2009). The first study confirmed a height association at the *Dw3* locus and subsequently fine-mapped a second major height locus to approximately 57 Mb on chromosome 9, using a panel of 378 sorghum inbreds that included 230 SC lines. It was demonstrated that the inclusion of the SC lines greatly increased the power of this study to detect height/flowering QTL, to the extent that significant height associations around the chromosome 9 QTL (*SbHT9.1*) extended for over 7 Mb, or approximately 1% of the sorghum genome. The second study used a panel of 125 mostly sweet sorghum inbreds genotyped with ~400 markers to test for association with plant height and sugar content (Brix). Although this marker density was too low for a full genome scan, this study nonetheless detected significant height associations near both the *SbHT9.1* and the *Ma1*/*Dw2* genomic regions. Since the major loci for plant height and flowering time in sorghum have been recent targets of intense selection by breeders, hitchhiking effects have clearly extended the selected haplotypes such that the detection of associated traits through genome-wide scans should be eminently feasible.

4 Identification of Additional Loci

4.1 Additional Photoperiod-Sensitivity Loci: Ma5, Ma6, and Ma7

The *Ma1–4*/*Dw1–4* model was changed substantially by Rooney and Aydin (1999), who reported the production of an extremely late-flowering hybrid from a cross between two early-flowering inbreds. This phenomenon was observed repeatedly in

crosses involving EBA-3, a grain sorghum line from Argentina. Analysis of ten F_2 populations with EBA-3 as a common parent showed than in eight populations, the data fit a two-gene complimentary dominant model, in which at least one dominant allele at each locus is required for photoperiod-sensitivity (late flowering). This model, which predicts 9:7 segregation for photoperiod sensitivity–insensitivity in the F_2, was further validated by analysis of $F_{2:3}$ populations, backcrosses to both parents, and testcrosses of EBA-3 to the milo maturity standards. The two loci involved were dubbed *Ma5* and *Ma6*, such that the EBA-3 inbred is *ma5Ma6*, and most other grain sorghum lines are *Ma5ma6*.

Notably, two of the ten F_2 populations showed a different pattern of inheritance. First, a cross between EBA-3 and R9025 yielded exclusively early-flowering F_2 individuals, indicating that R9025 carries the recessive *ma5* allele. Second, a cross between EBA-3 and Lahoma (a sudangrass) yielded only ~37% late-flowering individuals, significantly fewer than the ~56% expected under the 9:7 segregation model. Perhaps the simplest explanation for this result is segregation distortion. The simplest possible alternative genetic explanation—a single additional locus for which Lahoma has the recessive allele and at least one copy of the dominant allele is necessary for photoperiod-sensitivity— is not likely true: this model still overestimates the number of photoperiod-sensitive progeny by a significant margin ($\chi^2 = 4.904$, $p = 0.03$).

The fine-mapping of *Ma5*, *Ma6*, and an additional locus (*Ma7*) has recently been reported (Mullet et al. 2010). In order to map *Ma6*, the authors crossed EBA-3 (*ma5Ma6;* early flowering) to ATx623 (*Ma5ma6;* early flowering) to generate a late-flowering F_1, and backcrossed the F_1 to ATx623. Since ATx623 is fixed for dominant *Ma5*, segregation for flowering time in these BC_1F_1 individuals is expected to segregate 1:1 and map to the *Ma6* locus. *Ma6* was initially mapped to a ~13 cM region from 39 to 43 Mb on chromosome 6, and subsequently narrowed to an interval containing just 20 predicted genes. One of these genes is homologous to *PRR7* from Arabidopsis (*Oryza PRR37*), a component of circadian clock in both monocots and dicots, and sequencing the corresponding cDNA in EBA-3 and RTx430 (a *ma6* line) identified a nonsynonymous amino acid polymorphism in a proposed dimerization domain. Therefore, although the identity of *Ma6* has not been proven unequivocally, the sorghum *PRR7*/*PRR37* homolog is a strong candidate.

In order to map the *Ma5* locus, the authors crossed EBA-3 to A3Tx436 (*Ma5ma6*) to generate a late-flowering F_1, backcrossed the F_1 to EBA-3, and scored flowering time on ~3,000 BC_1F_1 individuals. Since EBA-3 is fixed for dominant *Ma6*, flowering time segregation in the progeny is predicted to depend solely on whether the F_1 contributes a *Ma5* or a *ma5* allele, with an expected 1:1 ratio of early–late flowering. Instead, only ~25% of BC_1F_1 individuals were early flowering, indicating the probable segregation of a second gene necessary to confer photoperiod-insensitivity. This additional locus was named *Ma7*. *Ma5* and *Ma7* were fine-mapped by bulk-segregant analysis to chromosomes 2 and 1, respectively (Mullet et al. 2010). Candidates in the 250 kb *Ma5* interval include a *COP9FUS5* gene involved in photomorphogenesis and a Myb-domain transcription factor. Candidates in the 400 kb *Ma7* interval include the phytochrome C locus and two MADS-box genes.

Why was the presence of *Ma7* not detected by Rooney and Aydin (1999)? If present in one of the F_2 populations analyzed, segregation for this locus should have increased the proportion of photoperiod-sensitive plants from 9/16 (~56%) to 45/64 (~70%), which was not observed. The authors did measure flowering time on 254 individuals from the same (A3RTx436 × EBA-3) × EBA-3 backcross that later indicated the existence of *Ma7*, and obtained slightly more than expected late flowering plants (138/254), but the departure from a 1:1 ratio was insignificant. Therefore, one possibility is that *Ma7* effects were present in the study by Rooney and Aydin (1999), but the number of progeny evaluated was too small to detect these effects.

4.2 Interaction of Photoperiod with a Thermosensitivity Locus, T

Additional investigation into the genetic control of flowering time in sorghum has been carried out by researchers in Japan (Tarumoto et al. 2003, 2005). In the first study, the presence of a single dominant locus conferring lateness was detected in two separate biparental populations (75% of F_2s late; 50% of BC_1s late when backcrossed to the early parent), but this phenotypic effect disappeared completely if the minimum temperate dipped below 20°C (Tarumoto et al. 2003). The dominant thermosensitivity locus was named *T*. Cold-induced acceleration of flowering in accessions with a dominant *T* allele was observed in short as well as long days. Tarumoto et al. (2005) analyzed two biparental populations with a flowering behavior similar to that reported by Rooney and Aydin (1999) for EBA-3 populations, wherein two relatively early parents give rise to a photoperiod-sensitive F_1 that flowers very late or not at all under long days. This phenomenon was originally attributed to the action of a photoperiod-sensitivity gene, *D*, acting epistatically with the thermosensitivity locus, *T*, in a manner similar to the interaction between *Ma5* and *Ma6*. Similarly to the situation observed during the fine-mapping of *Ma5*, however, Tarumoto et al. (2005) observed that only ~25% of BC_1F_1 individuals were early flowering, indicating the continued segregation of two loci in a backcross population where one allele of the dominant *D* locus was always present. The *D* locus was therefore split into two loci, D_1 and D_2. They also established that the critical daylength for the D_1–D_2-mediated photoperiod-sensitivity response was affected by the *T* locus, with a critical daylength between 12′15″ and 12′30″ in a *T*/- background with minimum growing temperatures over 20°C, and around 13′ in a *tt* background at any temperature or in a *T*/- background with minimum temperatures below 20°C.

One possible explanation for the relationship between the flowering loci reported from Texas and Japan is that the D_1, D_2, and *T* loci correspond to *Ma5*, *Ma6*, and *Ma7*. In BC_1F_1 populations with two of the three loci segregating, researchers in Texas reported an approximate 1:3 segregation of early–late plants (Mullet et al. 2010), whereas researchers in Japan observed a 1:2:1 segregation of early–middle–late (Tarumoto et al. 2005). The difference in latitude between Texas and Japan may account for this slight difference in the ability to discriminate different genotypic classes.

The failure to detect the segregation of *Ma7* in the initial experiment by Rooney and Aydin (1999) might also be explained by minimum temperatures below 20°C masking the effects of the *T* allele.

5 Bridging Classical and Molecular Genetics

Now that molecular data are cheap and plentiful, the bottleneck to understanding the genetic control of crop phenotypes clearly lies in germplasm development and the development of improved phenotyping methods. Fortunately, the accumulated efforts of sorghum breeders and geneticists over the past century have left us in a good position from which to move forward. Selection of spontaneous mutants during the early, hand-harvested years of US grain sorghum production has left us with multiple alleles at many of the major *Ma* and *Dw* loci (Table 14.2), which will prove invaluable for the identification of the underlying genes/polymorphisms. The years of effort invested in the Sorghum Conversion Program will yield further dividends when molecular studies identify additional loci—and additional alleles at known loci—that affect plant stature and maturity. Some of the major unanswered questions raised by the classical genetic studies (i.e.: the overdominance of *Ma1* in *ma2* backgrounds; the dependency of the effect direction of *ma2* on the allelic state at *Ma1*) may also finally be resolved using the power of modern genomics.

Identifying the major genes that control plant stature and maturity variation in sorghum is of fundamental interest to researchers outside the sorghum community. Cloning sorghum QTL for agronomically important traits will not only provide a logical starting point for reverse genetic efforts in *Saccharum* and *Miscanthus*, but may also shed light on the genetic control of the same traits in sorghum's more famous phylogenetic neighbor, maize. Although the maize genome has much higher diversity than the sorghum genome, making it easier to find and track polymorphism in genetic experiments, this same feature arguably makes problems of genetic heterogeneity more likely in maize than in sorghum. Also, the greater size and complexity of the maize genome may render it more prone to rearrangements that complicate mapping.

As more height and flowering QTL are cloned in sorghum, this will facilitate the identification of additional loci, and additional alleles of known loci, by calculating the phenotypic residuals from a model including all known causal polymorphisms. For example, SC lines that are unusually tall given their genotypes at *Dw3* and *SbHT9.1* likely harbor uncloned dwarfing polymorphisms; similarly, SC lines that are unusually short, given their genotypes, could contain independent mutations in *Dw3* or *SbHT9.1*.

The 8-locus classical model for the genetic architecture of plant height and flowering time in sorghum is unusually simple. It appears to be relatively accurate for temperate-grown grain sorghum, but is it accurate for sorghum in general? Given the current interest in sorghum and its relatives as bioenergy crops, the model should be extended to account for all types of sorghum, including sweet, forage, grain, and

bioenergy varieties. Included here are just a few of many specific questions that should be answered during the second century of sorghum genetics:

1. What genes determine height/size in photoperiod-sensitive types grown at temperate latitudes where they will never undergo floral initiation?
2. What is the contribution of allelic series at known *Ma* loci to the adaptation of sorghum landraces to different environments, particularly those in tropical latitudes and those with seasonal moisture fluctuations?
3. What effect, if any, do the *Dw* loci have on biomass composition?
4. Does molecular analysis of the SC lines provide evidence for additional *Ma/Dw* loci?
5. How does genetic background affect the penetrance of the *Ma/Dw* loci?

References

Brown PJ, Klein PE, Bortiri E, Acharya CB, Rooney WL, Kresovich S (2006) Inheritance of inflorescence architecture in sorghum. Theor Appl Genet 113:931–942

Brown PJ, Rooney WL, Franks C, Kresovich S (2008) Efficient mapping of plant height quantitative trait loci in a sorghum association population with introgressed dwarfing genes. Genetics 180:629–637

Childs KL, Cordonnier-Pratt MM, Pratt LH, Morgan PW (1992) Genetic regulation of development in *Sorghum bicolor*. VII. *ma3* Flowering mutant lacks a phytochrome that predominates in green tissue. Plant Physiol 99:765–770

Childs KL, Lu JL, Mullet JE, Morgan PW (1995) Genetic regulation of development in *Sorghum bicolor*. X. Greatly attenuated photoperiod sensitivity in a phytochrome-deficient sorghum possessing a biological clock but lacking a red light-high irradiance response. Plant Physiol 108:345–351

Childs KL, Miller FR, Cordonnier-Pratt M-M, Pratt LH, Morgan PW, Mullet JE (1997) The sorghum photoperiod-sensitivity gene, *Ma3*, encodes a phytochrome B. Plant Physiol 113: 611–619

Dahlberg J, Rosenow DT, Peterson GC, Clark LE, Miller FR, Sotomayor-Rios A, Hamburger AJ, Madera-Torres P, Quiles-Belen A, Woodfin CA (1998) Registration of 40 converted sorghum germplasms. Crop Sci 38:564–565

Harper J (1977) Plant population biology. Academic, London

Hart GE, Schertz KF, Peng Y, Syed NH (2001) Genetic mapping of *Sorghum bicolor* (L.) Moench QTLs that control variation in tillering and other morphological characters. Theor Appl Genet 103:1232–1242

Klein RR, Rodriguez-Herrera R, Schlueter JA, Klein PE, Yu ZH, Rooney WL (2001) Identification of genomic regions that affect grain mold incidence and other traits of agronomic importance in sorghum. Theor Appl Genet 102:307–319

Klein RR, Mullet JE, Jordan DR, Miller FR, Rooney WL, Menz MA, Franks CD, Klein PE (2008) The effect of tropical sorghum conversion and inbred development on genome diversity as revealed by high-resolution genotyping. Plant Genome 1:12–26

Lin Y-R, Schertz KF, Paterson AH (1995) Comparative analysis of QTLs affecting plant height and maturity across the Poaceae, in reference to an interspecific sorghum population. Genetics 141:391–411

Maunder AB (1999) History of cultivar development in the United States: from "Memoirs of A.B. Maunder—sorghum breeder". In: Smith CW, Frederiksen RA (eds) Sorghum: origin, history, technology, and production. Wiley, New York

Mullet JE, Rooney WL, Klein PE, Morishige D, Murphy R, Brady JA (2010) Discovery and utilization of sorghum genes (Ma5/Ma6). US Patent 20100024065, Jan 28 2010

Multani DS, Briggs SP, Chamberlin MA, Blakeslee JJ, Murphy AS, Johal GS (2003) Loss of an MDR transporter in compact stalks of maize *br2* and sorghum *dw3* mutants. Science 302:81–84

Murray SC, Sharma A, Rooney WL, Klein PE, Mullet JE, Mitchell SE, Kresovich S (2008) Genetic improvement of sorghum as a biofuel feedstock: IQTL for stem sugar and grain nonstructural carbohydrates. Crop Sci 48:2165–2179

Murray SC, Rooney WL, Hamblin MT, Mitchell SE, Kresovich S (2009) Sweet sorghum genetic diversity and association mapping for brix and height. Plant Genome 2:48–62

Pereira MG, Lee M (1995) Identification of genomic regions affecting plant height in sorghum and maize. Theor Appl Genet 90:380–388

Quinby JR (1974) Sorghum improvement and the genetics of growth. Texas A&M University Press, College Station

Rooney WL, Aydin S (1999) Genetic control of a photoperiod-sensitive response in *Sorghum bicolor* (L.) Moench. Crop Sci 39:397–400

Rosenow DT, Clark LE (1987) Utilization of exotic germplasm in breeding for yield stability. In: Fifteenth Biennial Grain Sorghum Research and Utilization Conference, 49–56

Rosenow DT, Dahlberg JA, Peterson GC, Clark LE, Miller FR, Sotomayor-Rios A, Queles-Belen A, Madera P, Woodfin CA (1996) Registration of fifty converted sorghums from the sorghum conversion program. Crop Sci 37:1397–1398

Rosenow DT, Dahlberg JA, Stephens JC, Miller FR, Barnes DK, Peterson GC, Johnson JW, Schertz KF (1997) Registration of 63 converted sorghum germplasm lines from the sorghum conversion program. Crop Sci 37:1399–1400

Schertz KF (1973) Single height-gene effects in hybrids of doubled haploid *Sorghum bicolor* (L.) Moench. Crop Sci 13:421–423

Sindhu A, Langewisch T, Olek A, Multani DS, McCann MC, Vermerris W, Carpita NC, Johal G (2007) Maize Brittle stalk2 encodes a COBRA-like protein expressed in early organ development but required for tissue flexibility at maturity. Plant Physiol 145:1444–1459

Stephens JC, Miller FR, Rosenow DT (1967) Conversion of alien sorghums to early combine genotypes. Crop Sci 7:396

Tarumoto I, Yanase M, Iwahara Y, Kuzumi Y, Morikawa T, Kasuga S (2003) Inheritance of a thermo-sensitivity gene controlling flower initiation in sorghum. Breed Sci 53:353–357

Tarumoto I, Yanase M, Kadowaki H, Yamada T, Kasuga S (2005) Inheritance of photoperiod-sensitivity genes controlling flower initiation in sorghum, *Sorghum bicolor* Moench. Jap Soc Grassland Sci 51:55–61

Chapter 15
Bridging Classical and Molecular Genetics of Sorghum Disease Resistance

Clint W. Magill

Abstract With rare exceptions, sorghum breeders have been highly successful in producing disease-resistant sorghums. Classical breeding techniques, including identification of resistant cultivars from large and readily available germplasm collections, crosses to elite but susceptible varieties, followed by backcrosses, selfing, and selection, have been used to develop disease-resistant breeding and commercial lines. However, for almost every disease, the breeding efforts must be continuous as new forms of the pathogens rapidly evolve to overcome resistance. Molecular tools that allow rapid and accurate tagging and identification of resistance genes permit, at the very least, the ability to use marker-assisted selection to combine different genes and to test the theory that stacking different genes for resistance will provide stable resistance. Knowledge of the nature and molecular functions of resistance genes promises much more: the ability to manipulate, alter, and enhance the signal transduction pathways that actually trigger host plant resistance.

Keywords Genetic resistance • Sorghum pathogens • Disease screening • Germplasm • Germplasm screening • Cytoplasmic male sterility • NBS-LRR resistance genes • Host defense responses • Durable resistance • Gene stacking • Quantitative trait loci

1 Background

As is true with all crop species, sorghum has its share of pathogens that can limit grain or forage production. While both bacterial and viral diseases occur wherever sorghum is grown, fungi cause most of the diseases that are reported. As will be

C.W. Magill (✉)
Department of Plant Pathology and Microbiology, Texas A&M University,
Room 202H LF Peterson Building 2132 TAMU, College Station, TX 77843-2132, USA
e-mail: c-magill@tamu.edu

A.H. Paterson (ed.), *Genomics of the Saccharinae*, Plant Genetics and Genomics: Crops and Models 11, DOI 10.1007/978-1-4419-5947-8_15,

described, an oömycete and a parasitic weed are also pathogens of sorghum, whereas there are conflicting reports concerning nematodes.

The manual "Descriptors for Sorghum" published jointly by the International Board for Plant Genetics Resources and the International Crops Research Institute for the Semi Arid Tropics requests all donors of accessions to the sorghum germplasm collection to provide information, when available, on susceptibility to 21 insects, 19 fungi, 3 bacteria, 5 viruses, and 3 Striga (parasitic weed) species. Although many nematode species have been reported to reproduce sufficiently on sorghum to reduce yields (Kollo 2002), the request does not mention nematodes. However, nematodes are included in the most recent Compendium of Sorghum Diseases (Frederiksen and Odvody 2000), a compilation that also includes description of 6 bacterial diseases, 8 viruses, approximately 29 fungal diseases, and 7 virus or MLO diseases. Examples of resistance breeding for each category of pathogens are described in Sect. 3.

1.1 Sources of Genetic Resistance

Fortunately, large numbers of sorghum accessions representing different species, subspecies, races, and hybrids provide a great range of phenotypic and genetic diversity that is available for screening. Entries have been collected from locations around the globe, especially those countries in Africa believed to be near the center of origin of cultivated sorghums (De Wet et al. 1970). A collection maintained through the auspices of the USDA has over 44,000 entries catalogued in GRIN http://www.ars-grin.gov/cgi-bin/npgs/html/desclist.pl?69, and ICRISAT also maintains more than 37,000 accessions with duplicate seed samples held in cold storage (see http://www.croptrust.org/documents/cropstrategies/sorghum.pdf). The USDA collection is maintained in Puerto Rico to permit year-round growth and avoid problems related to photoperiod-sensitive flowering. Because sorghum is naturally self-fertilizing, accessions are generally considered to be relatively pure lines. However, since outcrossing does occur, bagging of heads before pollen shed is used to assure the purity of specific lines. For the most part, interspecific hybrids are fertile, meaning that useful genes identified in one source can be transferred to another without resorting to the use of recombinant DNA technology. In fact, the ability to transfer genes through pollination also limits the prospects for use of genetic engineering in sorghum for traits such as herbicide resistance; any gene introduced to sorghum could easily move into Johnson grass (*Sorghum halepense*), which in the opinion of many, is one of the world's worst weeds.

1.2 Screens for Resistance

When a new pathogen or a new race of an old pathogen appears, identification of sources of resistance is generally the first step in resistance breeding programs. Because pathogens and their hosts have coevolved, with selection for resistance

favoring the host countered by selection for virulence genes favoring the pathogen (or vice versa), searches for resistance often start with accessions from areas where the host/pathogen association has been known to occur. For example, a screen of 109 accessions from 17 sorghum species showed that many weedy and wild accessions were resistant to a specific isolate of *Peronosclerospora sorghi*, the causal agent of sorghum downy mildew. The fact that many of these accessions originated from northeast Africa was not surprising, since the disease is common there. However, the observation that accessions originating in Australia were also resistant was unexpected, since sorghum downy mildew is not known to be present there (Kamala et al. 2002) and the pathogen is subject to strict quarantines.

In screening accessions in the USDA national collection, Erpelding (2008) identified lines from Ethiopia, Mali, Mozambique, and Sudan with resistance to anthracnose. In particular, the lines from Mali that came from regions with the highest rainfall where the disease is the most serious problem were observed to be most frequently resistant, suggesting that selection, natural or human, has already played a role.

INTSORMIL, the USAID-funded International Sorghum and Millet Collaborative Research Support Program, has routinely cooperated to screen promising new breeding lines being developed in participating state and national sorghum breeding programs. Entries for the INTSORMIL "all diseases and insects nursery" or ADIN are selected annually and grown in cooperator test plots at numerous locations worldwide. The lines are typically the most promising material from breeding programs. The ADIN allows investigators to determine field-level resistance to the prevailing pathogens and strains in their location and environment (the lines are usually grown in plots that are not inoculated). Standard checks (susceptible varieties) are included to allow assessment of disease pressure. To help protect intellectual property rights, cooperators have agreed not to provide seeds for direct use by commercial seed operations. Lines that are ready for formal release are included in a subset of material, the International Disease and Insect Nursery (IDIN) that is also made available to private breeders. Although the number of entries in the ADIN is limited annually to 70, in the past the entries included many converted lines (see below) and these have served as sources of broad-based disease resistance.

1.3 Data Available

Currently, disease ratings for ten diseases are included for various subsets of the 44,000 sorghum accessions listed in GRIN. In the case of downy mildew, separate ratings are available for a newly discovered race of *Peronosclerospora sorghi*, the causal agent that is also resistant to metalaxyl seed treatment. Dr. Hari Upadhyaya, the curator of the ICRISAT collection, has assembled a "minicore" collection of 242 sorghum accessions designed to include as much of the natural genetic variability as possible. That collection is currently being screened for anthracnose, ergot, head smut, and downy mildew resistance against isolates of the causal pathogens common to the USA.

The lines are also being characterized using AFLP and SSR DNA fingerprinting, which should help to identify those lines that differ in the basis for genetic resistance (Ramasamy et al. 2012).

2 Classical Resistance Breeding Schemes

Methods of sorghum hand emasculation and cross-pollination date back to 1918 (Nafziger 1918). Since hand emasculation is tedious and limiting as to the number of crosses that can be made, alternative methods of preventing pollen viability have been devised, including dipping the panicle for 10 min into 48°C water or covering the head with a plastic bag to accumulate sufficient moisture that released pollen will rupture. These alternatives have not been widely adopted, in part because they are less than 100% efficient, meaning marker genes need to be included to verify that any progeny are the result of a cross rather than inadvertent self-pollination.

Following the discovery of a cytoplasmic male sterility (CMS) system with nuclear suppressor genes in 1954 (Stephens and Holland 1954), the use of CMS to make crosses evolved rapidly and soon led to commercial production of high-yielding hybrids. The first CMS-restorer gene system came when a milo female was backcrossed extensively to a kafir male, giving rise to what is now designated as the A1 CMS system. Fertility can be restored by introducing a dominant restorer gene, *Rf1*. To avoid the experience with T cytoplasm in *Zea mays* that was associated not only with male sterility but also with susceptibility to a leaf blight pathogen, sorghum researchers have identified a number of other CMS-RF systems which are known to differ in mitochondrial defects and restorers (Schertz and Ritchey 1978; Xu et al. 1995). The standard breeding approach using CMS calls for the production of two isogenic lines, an A line which is male sterile because it has the male sterile cytoplasm and a maintainer B line which has fertile cytoplasm but no restorer genes. The R or restorer lines are then bred separately to provide both restorer genes and hybrid vigor to hybrid seed sold to growers. For single gene types of resistance, dominant genes can be provided by either parent but both parents (A/B and R) must be bred specifically to include required recessive resistance alleles.

Another boon to breeders, especially those in the USA and other temperate climates interested in grain sorghum, has been the "sorghum conversion program." Initiated as a joint Texas A&M University and USDA project in 1963, "alien" accessions collected from around the world are crossed to the same early-maturing, dwarf parent so that homozygosity for up to four maturity (Ma_1 to Ma_4) and four height genes (Dw_1 to Dw_4) can be selected in progeny to provide plants that can be harvested by combine (Stephens et al. 1967). This program has proven extremely successful in providing useful genes from exotic germplasm to be incorporated into today's breeding material and hybrids. However, as would be expected, regions around the selected *ma* and *dw* alleles do retain genomic blocks from the original common

donor as demonstrated both in an interspecific cross population (Lin et al. 1995) and *Sorghum bicolor* germplasm (Klein et al. 2008). Whereas early breeding efforts in the USA were based on a very narrow sample of the available sorghum germplasm, the sorghum conversion program has made genes from approximately 850 exotic cultivars available for incorporation into lines useful in temperate climates. These converted lines include representatives of all races, working groups, and locations of origin, potentially providing alternate genetic sources for key traits including insect and disease resistance. To quote from the 1996 US National Academy of Sciences report entitled *Lost Crops of Africa*: "All in all, the Sorghum Conversion Program has become one of the most successful plant-breeding programs ever; a model of achievement for crop scientists everywhere and with every crop. It provides populations that are reservoirs of genes, rather than a single, highly inbred variety." (National Research Council 1996).

As an example of the use of these programs, five (SC 326-6, SC 414-12E, BTX 378, B35-6, and TX 2862) of 50 accessions tested over 2 years were identified as resistant to the bacterium *Pseudomonas andropogonis* that causes bacterial leaf stripe of sorghum (Muriithi and Claflin 1997). The accessions designated SC originated from the sorghum conversion program and those beginning with B are the male fertile isoline of a male sterile "A" line.

An early reported success in breeding for disease resistance was against a disease at the time simply called milo disease. Milo disease was thought to be caused by a mix of pathogens including a Pythium species, but symptoms occurred only under favorable environmental conditions. Currently, the disease, which has only been reported in the USA, is attributed to *Periconia circinata* and is referred to as Periconia root rot. Melchers and Lowe (Melchers and Lowe 1943) recommended that instead of making crosses, it was simplest just to plant large numbers of bulked seeds of the desired susceptible variety in infested soils in greenhouse tests where any survivors will be resistant. Since sorghum typically shows approximately 5% outcrossing, it can be assumed that the recessive gene that conditions resistance was generally present in the population at some low level and homozygous in a few plants. As an example, only 20 of 12,500 seedlings of "Colby" survived, giving rise to resistant Colby.

More typically, once sources of resistance have been identified, a cross is made to adapted but susceptible cultivars and the gene(s) transferred through a series of backcrosses and selection. Most often this has led to deployment of single gene resistance, which in many cases can be overcome by mutations and/or selection for a new race (pathotype) in the pathogen population. An example with the disease anthracnose was the release of a resistant variety in 1965 that was no longer resistant just 3 years later (Reyes et al. 1969). Consequently, there has been a move toward breeding for "field resistance" or disease tolerance, which is generally assumed to be a quantitative trait with multiple contributing genes. Identification and tagging of quantitative trait loci or QTLs are feasible with modern DNA technology, as is described in a later section.

2.1 Multiple Traits

Breeders typically develop lines for varieties to be released for growth in a specific region that are resistant to prevailing diseases, even if the primary target of the breeding effort is not disease resistance. For example, lines selected for drought resistance at Kansas State University were also monitored for stalk rot resistance caused by three Fusarium species and *Macrophomina phaseolina*, fungi that are primarily responsible for the stalk rots seen there. Only SC599 showed relative resistance to both Fusarium and Macrophomina; SC134 gave small lesions following Fusarium inoculation and SC35 with Macrophomina, an indication that different genes are involved in response to these stalk rot pathogens (Tesso et al. 2005). It was also noted that most of the resistant accessions originated from Ethiopia or neighboring Sudan, where charcoal rot (caused by *Macrophomina phaseolina*) is endemic. Charcoal rot generally does not cause serious losses unless there are drought conditions, specifically at the time of grain development. Consequently, Diourte et al. (2007) screened for resistance in inoculated plants subjected to post-flowering water stress and were able to identify two cultivars (B35-6 and SC265-14E) that showed less charcoal rot damage than other varieties. They also identified a strain of the pathogen collected from a line previously considered resistant that caused more damage, indicating that races of *Macrophomina phaseolina* are present. Perhaps this accounts for the statement in a recent University of Missouri Extension publication (Wrather and Sweets 2009) that no resistant varieties are available. Because they cause lodging, it can be anticipated that control of these stalk rot pathogens will become of critical importance in the extremely tall varieties being developed for production of biomass to serve as an energy source.

Texas germplasm releases over the past several years are generally noted for response to diseases that are especially important in the target growing area. For example 16 lines released in 2006 were targeted specifically for resistance to biotype E greenbug [*Schizaphis graminum* (Rondani)]. By also starting with a disease-resistant parent and selecting progeny rows in areas of high disease incidence, only lines that all were also resistant to head smut [*Sphacelotheca reiliana* (Kuhn) Clinton], rust [*Puccinia purpurea* (Cooke)], zonate leaf spot [*Gloeocercospora sorghi* (Bain and Edgerton)], bacterial leaf streak [*Xanthomonas holcicola* (Elliot) Starr and Burkholder], and bacterial leaf stripe [*Pseudomonas andropogoni* (E.F. Smith) Stapp] were released (Peterson et al. 2009).

3 Information on Specific Pathogens

3.1 Viruses

Although sorghum is considered an alternate host, Maize Dwarf Mosaic Virus (*MDMV*) is the most prevalent virus for which sorghum infections are noted. Fortunately, the virus is not seed transmitted but requires aphids, greenbugs, or

mechanical means of introduction (Toler 1985). By screening large numbers of cultivars, both tolerant and fully resistant cultivars have been identified. Full resistance is attributed to a single dominant gene referred to as the Krish gene based on its identification in "Krish" sorghum, a *Sorghum halepense* cultivar from India. The Krish gene has been introduced to *Sorghum bicolor* via formation of interspecific hybrids followed by backcrossing and selection into breeding lines used for commercial hybrids in Australia. The resistance seems to be stable and has been effective in preventing MDMV from being a truly serious problem. One of the Australian lines (QL3) was converted and adapted for use in the USA (Toler 1985). Seed companies such as Garst Seed Co. and Sorghum Partners generally provide information on MDMV resistance or tolerance in the commercial hybrids currently being sold. Although multiple strains are known, only MDMVA and MDMVB are usually associated with sorghum. Unlike the A strain of MDMV, the B strain does not overwinter in Johnson grass rhizomes. The same Krish gene or a very closely linked gene also provides tolerance to sugarcane mosaic virus (Condé et al. 1976).

3.2 Bacteria

Extension publications from several US sorghum-growing states indicate that bacterial leaf stripe and *bacterial leaf streak* caused by *Pseudomonas andropogonis* (now generally called *Burkholderia andropogonis*) and *Xanthomonas holcicola*, respectively, are the predominant local bacterial diseases, but also indicate that they generally do not cause severe loss. These statements most likely refer to grain sorghums and if sorghum biomass becomes a key source of nonfossil energy, leaf diseases may rise to higher importance. Examples of screens that have identified resistant cultivars can be seen for *Pseudomonas andropogonis* (Muriithi and Claflin 1997 and for *Xanthomonas holcicola* and *Pseudomonas andropogoni*) (Peterson et al. 2009).

3.3 Nematodes

Worldwide, losses of grain sorghum to *nematodes* in 1987 were estimated at 6.9% compared to 15.3% for sugarcane, both of which are much higher than the US loss estimates (Koenning et al. 1999). To point out the differences that climate, locale, or perhaps host cultivars can make, *Meloidogyne incognita*, the cotton root-knot nematode, was found to reproduce very effectively on all 23 sorghum cultivars tested in an Arizona test (MacClure et al. 1998) where cotton yields were noticeably lower after the field had been planted with sorghum. On the other hand, a set of commercial hybrids tested in South Carolina did not support reproduction of nematodes at

greater than replacement rates (Fortnum and Currin 1988). In fact, sorghum or sorghum cross-species hybrids are recommended as part of a crop rotation to lower nematode populations in southeastern US soils (McSorley et al. 2004).

3.4 Striga

The parasitic weed striga may well be the most serious pathogen of sorghum in the areas where sorghum is an important food crop. Work done over many years by 2009 World Food Prize recipient, Dr. Gebesa Ejeta, has revealed several avenues to approach resistance breeding. Laboratory screens were first developed that targeted critical areas of host–pathogen signaling required for striga seeds to germinate and attach to sorghum roots. Screening a large collection of cultivated sorghums as well as a smaller group of wild sorghum accessions identified several different genetic variants, including some unique to the wild sorghum accessions (Rich et al. 2004). Specific striga resistance mechanisms described include resistance associated with low germination stimulant (LGS) production, low production of the haustorial initiation factor (LHF), hypersensitive response (HR), and a response incompatible (IR) to parasitic invasion of host genotypes (Ejeta 2007). Single genes for LGF (recessive) and LHF (dominant) were identified, while either of the two dominant genes was found to trigger HR in CK32 and KP33. Crosses made between cultivars with different modes of defense have also been used to loosely tag QTL that contribute to LGF, LHF, and HR in selected crosses. The Ejeta group has made crosses that will permit combining alternate sources of resistance into broadly adapted cultivars and for introgression of the traits into locally selected African landraces. As of 2007, several releases of striga-resistant cultivars were being field tested in Africa.

In another approach, Tuinstra et al. (2009) have identified a natural mutation that makes sorghum tolerant to acetolactase-inhibiting herbicides, thus permitting the use of these herbicides as a seed treatment to reduce and delay the emergence of Striga.

3.5 Oömycetes

Although oömycetes resemble fungi and in fact have been classified as such for years, molecular data indicate that they are more closely related to algae, leading to Oömycota being reclassified as a separate group. *Peronosclerospora sorghi*, an oömycete that is the causal agent of *sorghum downy mildew*, is found wherever sorghum is grown with the exception of Australia. Downy mildew has traditionally been controlled by a combination of resistant varieties and seed treatment with phenylamide-type systemic fungicides such as metalaxyl. In recent years both metalaxyl resistance and a new race that reproduces on previously resistant hybrids have been reported (Isakeit and Jaster 2005). The need to identify new sources of resistance led to tests of accessions in the USDA collection from Chad (78 accessions tested) and Uganda (20 accessions tested). Those tests were successful in identifying six accessions that were resistant when

grown in plots in the area where the new race was identified, so the basis for resistance is expected to differ from that of prior sources (Prom et al. 2010). Similarly 6 of 40 accessions from China were found to be resistant (Prom et al. 2007).

As is often the case, close relatives of *P. sorghi* such as *P. maydis* and *P. sacchari* can also infect sorghum, though as the names imply they have alternate host preferences. One species, *P. philippinensis* (formerly *Sclerospora philippinensis*), is especially virulent to sorghum, maize, and sugarcane. Because *P. philippinensis* has been reported only in the Philippines and Thailand, strict quarantine measures are employed to try to prevent its spread. In fact, *P. philippinensis* has been considered a potential "biothreat" to maize production, so has been placed on the US list of select agents. For maize, resistant cultivars were developed in Thailand by making selections from a random mating population of 36 varieties. The original composite variety, Suwan 1, has subsequently served as a source of downy mildew resistance in maize, including the development of inbred lines (Sriwatanapongse et al. 1993). One of the possible avenues of future research is to determine if the genes that function to give maize broad resistance to downy mildew will also function in sorghum.

3.6 Fungi

In contrast to many pathogens, extensive screening was required to identify sorghum accessions with resistance to *sheath blight* caused by *Rhizoctonia solani*. Before identification of a single resistant line (CS 621) among thousands tested, leaf blight was a major problem in the Philippines, causing yield losses estimated at 35–75% (Pascual et al. 2000). Crosses to susceptible lines showed that both additive and dominant QTLs contributed to resistance. CS 621was also found to be a source of resistance to tar spot (*Phyllachora sorghi*) and grey leaf spot (*Cercospora sorghi*), but no attempt was made to see if the same QTLs were involved.

Resistance to *anthracnose* (caused by *Colletotrichum sublineolum*, but generally referred to as *C. graminicola* in papers before 1999 and sometimes since) is typical; new races of the pathogen mean that new sources of resistance must be identified and incorporated into breeding lines. Starting with crosses between 11 resistant accessions and using the same pathogenic strain for inoculations, Mehta et al. were able to identify five complementing groups meaning that at least five different genes can lead to resistance. With one possible exception, resistance was inherited through single genes, at least three of which are dominant. Follow-up field tests in five locations (three in the USA and two in West Africa) revealed only one source that maintained resistance across all locations, a testimony to the variability of the pathogen (Mehta et al. 2005). A complicating factor is exemplified in a genetic analysis of anthracnose resistance in the variety Redlan. F2 Progeny of a cross to a highly susceptible parent showed that a single dominant gene controlled resistance to foliar anthracnose, but that an unlinked recessive gene controlled infection of the midrib (Erpelding 2007).

Rust resistance was investigated by Tao et al. (1998) using 166 segregating restriction fragment length polymorphism (RFLP) markers. They were able to

identify four QTLs segregating for rust (*Puccinia purpurea*) resistance in a recombinant inbred population derived from a cross between QL39 and QL41. Each of the genes mapped to a different chromosome, with 40% of the contribution to resistance associated with an allele from QL 41, the parent with greater resistance. However, transgressive segregation of the rust scores in the 160 RI lines indicated that different genes contributing to resistance are present in both parents. In what may become a precursor of future breeding possibilities, a resistance gene analog or RGA has been located in the same map region on linkage group 10.

Smut fungi are basidiomycetes that typically infect germinating plants from spores present in the soil, but do not produce symptoms until the time of flowering. Three species of *Sporisorium* (formerly *Sphacelotheca*) are common sorghum pathogens. *S. sorghi* causes covered kernel smut in which individual grains are replaced by sori, the reproductive body of the fungus. *S. cruenta* causes loose kernel smut which also replaces individual grains with sori but also disrupts formation of other flower parts. The most severe is *S. reiliana* in which the whole head may be replaced by a gall that is a mass of fungal reproductive tissues. Some strains of *S. reiliana* can also cause head smut of maize but based on electrophoretic karyotypic differences, these could perhaps more appropriately be considered a maize-specific *forma speciales* or separate subspecies (Naidoo et al. 1999). Some of the earliest reports of the genetic basis for disease resistance in sorghum deal with smuts (Marcy 1937; Swanson and Parker 1931) and even the earliest recognizes the existence of different forms of the pathogen that require different host genes for control. It was also noticed that some but not all genes are effective against both kernel smuts. Host differentials that can be used to define races of *S. reiliana* have been established, and a recent survey indicated the presence of two new races in samples collected from south Texas (Prom et al. 2011). Even though sorghum varieties resistant to different races are used to control the disease, the much greater economic impact of maize has led to more research on *S. reiliana* resistance than is available for sorghum. For example, 580 maize genotypes were screened in Kenya and at least four were identified as immune (0% infection when highly susceptible checks recorded 100%) (Njugama 2001). Other studies have identified four QTLs contributing to resistance in a cultivar grown in China where the disease is prevalent (Li et al. 2008). A QTL on maize chromosome 2 in that study accounted for over 40% of the variation seen in 184 segregating F2–F3 progeny.

Although *ergot* has been recognized as a problem in Asia and Africa for nearly a century, it only was introduced into the Americas and Australia in the mid 1990s (Bandyopadhyay et al. 1998). Rapid spread from South to North America led to considerable research to identify sources of resistance and control methods. An unexpected outcome of the research was the realization that two different species of the pathogen are extant; in Asia, *Claviceps sorghi* is the causal agent while in Africa, it is *Claviceps africana* and it is this species that has seen rapid global expansion. (Both are anomorphs of the ascomycete *Sphacelia sorghi*, but are classed as different species based on differences in alkaloid production and shape of the sclerotia they produce.) Although sorghum ergot has not led to the problems with neurotoxic

and abortion-inducing alkaloids produced by other Claviceps species, ergot infection still poses severe production and marketing problems to sorghum growers and hybrid seed producers. The primary problem is that ergot only infects through the stigmas of unfertilized ovaries, i.e., before pollination has occurred. Infection leads to the displacement of kernels with the production of massive amounts of conidia that are exuded from sphacelia in the infected florets, creating a sticky mass called honeydew that drips on the leaves and to the ground. The honeydew can cause combines to clog during harvest and prevent flow of grain from trucks, causing problems beyond the loss of grain. Of even greater concern to the hybrid seed industry is the fact that the male sterile lines used are especially susceptible due to the fact that their pistils are typically exposed for a much longer period of time before fertilization. As a consequence, massive efforts have been taken to discover sources of genetic resistance (Prom et al. 2008). Resistance has been associated with physiological traits such as production of copious pollen or short duration of open florets, traits that are not compatible with hybrid production through the use of male steriles. Although QTLs that contribute to ergot resistance have been identified (Parh et al. 2008), at least some of them contribute to physiological traits that may hinder direct application in breeding programs. Based on the mode of infection and the difficulty in identifying more typical resistance genes, it seems likely that the signal transduction pathways that function to activate host defenses in other tissues do not function in stigmas. The rationale for this suggestion is that host defenses would likely also be triggered by growth of pollen tubes through the style and result in activation of pathways potentially harmful to fertilization, including synthesis of reactive oxygen species and even apoptosis.

Grain mold is a good example of a trait that requires multiple genes for effective resistance. Many different fungi, of which *Curvularia lunata* and *Fusarium thapsinum* are the most common, can cause grain mold (Little and Magill 2009). Consequently, prospects for finding a single gene that will prevent grain mold are unrealistic. Further, grain mold is greatly affected by environmental conditions and of course by the presence of grain mold fungi, so heritability of resistance can be quite low, a situation that makes identification of contributing alleles especially difficult. However, cultivars grown in the same environment do differ greatly in the level of grain mold that develops, verifying that genes and gene combinations can have a major impact on the amount of molding seen in the field. This has been verified by QTL analysis of F5 recombinant inbred progeny from a cross between Sureno (relatively resistant) and RTx430 which is very susceptible. Alleles (all from Sureno) that contributed to mold resistance across several locations were found on four linkage groups, and on a fifth for one environment. These were different from other QTL for traits such as plant height and panicle shape, kernel hardness, and tannin content which also affect the extent of grain mold. The last two characters negatively impact digestibility, so finding alternative genes for resistance is critical for grain sorghums to be used as human or animal food.

4 Toward More Durable Resistance: Marker-Assisted Selection

4.1 Gene Stacking

It has long been assumed that if different genes that each provide resistance to a specific pathogen race can be combined, new pathogenic races would evolve less frequently. However, to combine multiple genes for resistance that have the very same phenotypic effect is extremely difficult in a classical breeding scheme. In order to ensure that both (of two) genes are present, especially if recessive, in an F2 or later progeny, each plant would have to be backcrossed to both parents and the segregation ratio in the progeny verified under high disease pressure with controls to counter the risk that escapes will be classified as resistant. To introgress a gene from an alien species with undesirable agronomic traits, multiple generations of breeding will be required to reduce linkage drag, and either very large populations must be maintained or every generation will require results to be verified from backcross progeny testing before proceeding. Then, if an additional gene is to be added, the process will have to be repeated with additional progeny tests. It is this difficulty and the fact that DNA polymorphisms are abundant that has inspired the development of molecular markers from highly polymorphic DNA sequences scattered throughout the genome and their use for marker-assisted selection (MAS).

A variety of techniques have been developed that are effective in identifying DNA-based polymorphisms between two cultivars. Rapid progress has been made since the original Southern blotting procedure of detecting of RFLPs, with a variety of polymerase chain reaction (PCR)-based systems that require much smaller DNA samples and in many cases allow automated detection (Bowers et al. 2003; Menz et al. 2002). In essence, the basis for tagging resistance or other genes is to detect co-segregation of a specific phenotype with a DNA-based polymorphism. Examples of the use of PCR to tag a single gene for resistance include the use of short random primers (RAPDs) for downy mildew (Oh et al. 1996) and a recessive gene for anthracnose (Boora et al. 1998). RAPDs and sequence characterized amplified regions (SCARs) were used to tag a different gene for anthracnose resistance (Singh et al. 2006). Simple sequence repeats (SSRs), which are also referred to in some cases as microsatellites, have been used for a leaf blight resistance gene (Mittal and Boora 2005) and both SSRs and amplified fragment length polymorphisms (AFLPs) were used to tag *Cg1*, a dominant gene for resistance to anthracnose (Ramasamy et al. 2009). The advantage of SSRs and some AFLP markers is that many have been mapped to specific chromosomes, so that other nearby potentially useful markers are identified along with the chromosomal location of the resistance gene. While none of the prior gene tags described are ideal in that they are several map units from their target, they all provide at least some sequence information that will make them more useful now that the sorghum genome sequence is available. Because

SSRs tend to be randomly distributed and highly polymorphic, and since they can be detected by the size of the amplicon produced by flanking PCR primers, they seem to be ideal for narrowing the gap between gene tags and their target, at least until the gene itself can be identified. Recent scans of the sorghum genome sequence have identified approximately 6,000 new SSR sites, half of which are expected to be useful in most crosses based on a sample of 970 that were tested (Yonemaru et al. 2009). Thus it would now be feasible to use MAS to combine two or more different tagged anthracnose resistance genes into individual cultivars and test the hypothesis that the combination will lead to stable resistance. In addition, as described below, combining gene tags with the genome sequence will promote discovery of the function and mechanisms of individual resistance genes.

4.2 Quantitative Resistance and QTLs

Many factors in addition to activation of host defense responses can slow the spread of disease, leading to what is often referred to as field or quantitative resistance. In general, quantitative resistance is predicted to be more durable than single gene resistance since there is less selective pressure against the pathogen; the reproductive rate is slowed and even races that are less virulent are still able to reproduce. Grain mold resistance for example is affected by panicle shape, grain hardness, antifungal proteins stored in the seed, polyphenols, and presence or absence of glumes along with active defense responses, all of which show genetic variation. Thus crosses between lines that differ in mold-preventing factors can lead to a redistribution of the genes encoding these traits, allowing selection for those lines with the best combination of contributing loci (QTLs). Furthermore, even if the nature of the factor contributing to resistance is not known, the contributing alleles can be mapped, tagged, and identified for potential use in other crosses. (Potential is used since it is possible that epistasis or other complications could alter expression in different genotypes.) Identification of key contributing alleles, along with the relative contribution each makes to overall variation, is generally initiated by identifying DNA-based polymorphisms that separate the parents and can be mapped to intervals along each of the ten sorghum chromosomes. The very large number of SSR loci that have now been mapped and the simplicity and accuracy of detecting polymorphisms will likely make these markers the choice for future studies. By measuring expression of each trait for each progeny in a segregating cross, it is possible to determine when co-segregation with any parental marker is associated with high or low expression (see Lin et al. (1995) for an example of QTL mapping of height genes in sorghum). Software applications for handling QTL mapping data have been developed for most computer platforms.

In some instances, especially if relatively few genes are contributing to resistance, it may be simpler to identify alleles of contributing resistance genes or QTLs through bulked-segregant analysis. In this procedure, DNA samples from individuals within each tail of the distribution are combined and amplified with primer pairs for each

potential SSR polymorphic site; appearance of a parental band in only one of the bulks indicates close linkage to a contributing allele. In this case, AFLPs could also be used. If an amplified band in the resulting fingerprint has not already been mapped, it is sometimes possible to sequence the band in order to provide information that should now allow its location to be determined. Note that even if the QTL has not been mapped, a closely linked polymorphism can still be used to track transfer of the gene, and thus to combine contributing alleles from one parent or both parents into a new cultivar with an improved level of tolerance to a pathogen.

Taking a third approach based on identification of shared characters in numerous diverse lines, Casa et al. 2008 have established a panel of 377 accessions (228 converted lines and 149 breeding lines) suitable for use in association mapping, where DNA-based polymorphisms such as SSRs can be used to identify common patterns associated with specific traits. In addition to helping identify loci that contribute to quantitative traits, the panel should aid in identifying genes for disease resistance that represent the same allele based on common descent as well as uniquely different genes that provide alternate pathways to resistance.

5 Understanding/Engineering Resistance

5.1 Defense Response Genes and Disease Resistance

When plants sense the presence of a potential pathogen, a series of genes are activated to combat and counter the threat. The responses are general and may be activated by fungi or bacteria including those that are not considered pathogens. Insects and environmental stresses can also trigger defense responses. Typical host response genes include enzymes such as chitinase that attack fungal cell walls, polygalacturonase-inhibiting proteins to inactivate fungal enzymes that degrade plant cell walls, glucosinolates that break down to toxic compounds, peroxidases that create toxic oxygen free radicals, caspaces that lead to apoptosis (natural cell death), and an array of phytoalexins that are host specific (Bednarek and Osbourn 2009). The sorghum phytoalexins are flavanoid compounds derived from phenylpropanoids and have been identified as luteolinidin, 5-methoxy-luteolinidin, apigeninidin, and the caffeic acid ester of arabinosyl 5-*O*-apigeninidin (Du et al. 2009; Nicholson et al. 1987). In general, defense responses are activated in both resistant and susceptible hosts, so the speed, level, or duration of the response can be critical in determining whether the plant or pathogen "wins out." Differential expression of defense response genes in sorghum following inoculation with a potential pathogen is most readily demonstrated by measuring levels of specific mRNAs, originally by northern hybridization (see Cui et al. 1996) or, more recently, by real-time reverse transcriptase PCR (see Katilé et al. 2009) for examples.

5.2 The Role of R Genes

The primary questions concerning host–pathogen interactions that lead either to disease or resistance are: "How do plants sense the presence of a pathogen and what is an R (resistance) gene?" Initial recognition likely involves membrane-spanning host proteins that initiate responses sufficient to repress non-pathogens. However, for the organisms that escape this level of response, especially biotrophic pathogens, resistance or R genes expressed inside host cells serve as a more effective recognition trigger and lead to rapid induction of active defense responses (Takken and Tameling 2009). Among the small sampling of R genes cloned, most have been found to encode proteins with conserved nucleotide binding site (NBS) and leucine-rich repeat (LRR) motifs. The LRR motif allows for interactions of the R gene product with other proteins, possibly of pathogen origin, and possibly in combination with host components (the guard hypothesis) to activate NB activity which serves to alter the R protein conformation (Takken and Tameling 2009). The new conformation provides the ability to initiate signal transduction pathways that in the end interact with enhancer elements in the promoters of defense response genes to activate expression. Extremely rapid response can trigger an HR in which host cells die to prevent the spread of a biotrophic pathogen. Significant progress is being made in determining the pathogen virulence factors and elicitors involved in triggering host defense responses, especially for bacterial systems. Bacterial proteins secreted into host cells are referred to as pathogen-associated molecular patterns (PAMPs) and include flagellin and EF-Tu, a protein required for translation of bacterial mRNA. In studies using Arabidopsis as a model organism, it has been demonstrated that each of these proteins is recognized by pattern recognition receptors (PRRs) to trigger initial host defenses (Zipfel et al. 2006). In the current model, successful bacterial pathogens then make other proteins that allow them to escape the early host defense response, and it is these new virulence factors that are recognized by specific R genes of the NBS-LRR class. While less information is available for fungal pathogens, it seems likely, especially for cases of "gene-for-gene" systems where a different R gene may be required for each of the races of a pathogen, that each R gene interacts with a different fungus-derived elicitor. The best examples come from the interaction between flax R genes and the corresponding *Avr* (avirulence) genes in the flax rust pathogen *Melamspori lini*. Successful interactions lead to effector-triggered immunity (Catanzariti et al. 2010).

While specific models have not been evolved in sorghum for any class of pathogen, it has been possible to use the genome sequence to identify sorghum R gene analogs. So far 211 have been identified (Paterson et al. 2009). In a sense this makes sorghum a better model system to work with than rice or Arabidopsis, both of which have many more R genes in their genomes. Work to tie tagged resistance genes to specific R genes in the sorghum genome has been initiated.

Elucidation of the genes and signal transduction pathways involved in host defense promises approaches for "molecular breeding" that go beyond the use of MAS for gene stacking. For example, an R gene from one host can be effective in

another: the *Rxo1* gene of maize that makes it effective against *Burkholderia andropogonis*, the causal agent of bacterial stripe in maize and sorghum, was shown to recognize *Xanthomonas oryzae* pv. *oryzicola*, a pathogen of rice and not of maize (Zhao et al. 2005). In contrast *Rp*1, another NBS-LRR gene, did not function when transferred to wheat or barley. As a long-term prospect, it should eventually be possible to identify specific elicitor/LRR interactions and use the information to combine R genes whose interactions complement each other. It may also be possible to create new R genes in vitro by site-directed mutagenesis or by insertion of random bases into an LRR DNA synthesis system and then selecting out those that encode proteins that interact with specific elicitors for amplification and reinsertion.

Another avenue that should be explored involves "stealth" pathogens such as the smuts that survive in the host for a long period of time without activating a defense response. Initial infection can be detected as demonstrated by the fact that smut-resistant cultivars have been identified and host differentials established for different races. However once the pathogen gains access, it essentially turns infected plants into weeds: at flowering fungal teliospores are produced rather than seed. Would it be possible to induce defense responses to counter the infection or at least prevent inoculating the soil with infectious smut propagules? In many systems host defense can be activated chemically, for example by salicylic acid or glufosinate ammonium, compounds that have been used to create systemic acquired resistance (SAR) (Ahn 2008; Vallad and Goodman 2004). SAR can help to prevent subsequent damage by pathogens but effects on systemic pathogens are yet to be determined.

5.3 Other Resistance Mechanisms

Not all effective R genes are based on NBS-LRR proteins. An excellent example is *Lr34*, a gene that has provided stable resistance to leaf rust, stripe rust, and powdery mildew to wheat cultivars for more than 50 years. Although its mode of action has not been established, the amino acid sequence shows that it is a kind of protein called an ABC transporter that is also involved in multiple drug resistance. One hypothesis for its function is that it prevents fungal takeover by removing key fungal metabolites from host cells, thus preventing disease (Krattinger et al. 2009). It is not known if an analogous gene is found in sorghum or if transfer of the wheat gene to sorghum would have any effect on disease resistance.

The best described system for defense against RNA viruses does not involve R genes, but may offer a mechanism that can be developed to protect against other organisms that directly feed on plants. Posttranscriptional Gene Silencing is triggered by the presence of dsRNA and likely evolved as a natural defense against RNA viruses. The same system was found to occur in *C. elegans*, a free-living nematode, when it was found that dsRNA was more effective in gene silencing than antisense RNA (Fire et al. 1998), and led to the understanding of what is now called RNA interference (RNAi). Cellular components found in almost all eukaryotic systems recognize dsRNA sequences, dice them into fragments, and then use the

fragments as a guide to destroy matching sequences of any mRNA (Sontheimer 2005). Because the trigger seems to be simply the presence of dsRNA, it has been possible to develop a system called virus-induced gene silencing (VIGS) whereby genetically modified virus can express for example a foldback copy (hairpin) of any gene to trigger the destruction of that gene's mRNA, creating a powerful tool for studying gene function (Benedito et al. 2004). In theory it should be possible to prevent damage from either DNA or RNA viruses in sorghum by transforming plants to constitutively express a transgene that would create a foldback when transcribed to target one or more key viral messages. However, it must be pointed out that successful viruses such as MDMV have already evolved a mechanism for avoiding PTGS. A similar system might also apply to nematodes that can take up small RNAs from host cells during feeding, at least for those like *C. elegans* where assimilated dsRNA fragments spread throughout the organism. In that case, expressing a hairpin copy of almost any gene critical to the life cycle of a nematode would provide resistance. Examples in Arabidopsis and tobacco model systems suggest that it would work for nematodes that attack those species (Sindhu et al. 2009; Yadav et al. 2006). Proposals that RNAi could be used to control striga and even fungal pathogens have been made, but thus far there has been no evidence that intact dsRNA expressed by the host is taken into the pathogen. This is one instance where the fear that genes used to create transgenic sorghums could be transferred to Johnson grass (Sangduen and Hanna 1984) through interspecific crosses might be less a factor than for genes such as herbicide resistance, unless lowered reproduction of nematodes on Johnson grass also greatly improved its relative fitness.

References

Ahn I-P (2008) Glufosinate ammonium-induced pathogen inhibition and defense responses culminate in disease protection in BAR-transgenic rice. Plant Physiol 146:213–227

Bandyopadhyay R, Frederickson DE, McLaren NW, Odvody GN, Ryley MJ (1998) Ergot: a new disease threat to sorghum in the Americas and Australia. Plant Dis 82:356–367

Bednarek P, Osbourn A (2009) Plant–microbe interactions: chemical diversity in plant defense. Science 324:746–748

Benedito V, Visser P, Angenent G, Krens F (2004) The potential of virus-induced gene silencing for speeding up functional characterization of plant genes. Genet Mol Res 3:323–341

Boora KS, Frederiksen R, Magill C (1998) DNA-based markers for a recessive gene conferring anthracnose resistance in sorghum. Crop Sci 38:1708–1709

Bowers JE, Abbey C, Anderson S, Chang C, Draye X, Hoppe AH, Jessup R, Lemke C, Lennington J, Li Z, Y-r L, S-c L, Luo L, Marler BS, Ming R, Mitchell SE, Qiang D, Reischmann K, Schulze SR, Skinner DN, Wang Y-W, Kresovich S, Schertz KF, Paterson AH (2003) A high-density genetic recombination map of sequence-tagged sites for sorghum, as a framework for comparative structural and evolutionary genomics of tropical grains and grasses. Genetics 165:367–386

Casa AM, Pressoir G, Brown PJ, Mitchell SE, Rooney WL, Tuinstra MR, Franks CD, Kresovich S (2008) Community resources and strategies for association mapping in sorghum. Crop Sci 48:30–40

Catanzariti A-M, Dodds PN, Ve T, Kobe B, Ellis JG, Staskawicz BJ (2010) The AvrM effector from flax rust has a structured C-terminal domain and interacts directly with the M resistance protein. Mol Plant Microbe Interact 23:49–57

Condé B, Moore R, Fletcher D, Taeackle D (1976) Inheritance of the resistance of Krish sorghum to sugarcame mosaic virus. Aust J Agric Res 27:45–52
Cui YC, Magill J, Frederiksen RA, Magill C (1996) Chalcone synthase and phenylalanine ammonia-lyase mRNA levels following exposure of sorghum seedlings to three fungal pathogens. Physiol Mol Plant Pathol 49:187–199
De Wet JMJ, Harlan JR, Price EG (1970) Origin of variability in the Spontanea complex of *Sorghum bicolor*. Am J Bot 57:704–707
Diourte M, Starr J, Jeger M, Stack J, Rosenow D (2007) Charcoal rot (*Macrophomina phaseolina*) resistance and the effects of water stress on disease development in sorghum. Plant Pathol 44:196–202
Du Y, Chu H, Wang M, Chu IK, Lo C (2009) Identification of flavone phytoalexins and a pathogen-inducible flavone synthase II gene (*SbFNS*II) in sorghum. J Exp Bot 61(4):983–994, Erp364
Ejeta G (2007) Breeding for Striga resistance in sorghum: exploitation of an intricate host parasite biology. Crop Sci 47:S-216–S-227
Erpelding J (2007) Inheritance of anthracnose resistance for the sorhgum cultivar Redlan. Plant Pathol J 6:187–190
Erpelding JE (2008) Sorghum germplasm resistance to anthracnose. Am J Plant Sci Biotechnol 2:42–46
Fire A, Xu S, Montgomery M, Kostas S, Driver S, Mello C (1998) Potent and specific genetic interference by double stranded RNA in *Caenorhabditis elegans*. Nature 391:806–811
Fortnum B, Currin R III (1988) Host suitability of grain sorghum cultivars to *Meloidogyne* spp. J Nematol 20:61–64
Frederiksen R, Odvody G (2000) Compendium of Sorghum diseases, 2nd edn. APS Press, St. Paul, Minnesota
Isakeit T, Jaster J (2005) Texas has a new pathotype of *Peronosclerospora sorghi*, the cause of sorghum downy mildew. Plant Dis 89:529
Kamala V, Singh SD, Bramel PJ, Rao DM (2002) Sources of resistance to downy mildew in wild and weedy sorghums. Crop Sci 42:1357–1360
Katilé S, Perumal R, Rooney W, Prom L, Magill C (2009) Expression of pathogenesis-related protein PR-10 in sorghum floral tissues in response to inocluation with *Fusarium thapsinum* and *Curvularia lunata*. Mol Plant Pathol 11:93–103
Klein RR, Mullet JE, Jordan DR, Miller FR, Rooney WL, Menz MA, Franks CD, Klein PE (2008) The effect of tropical sorghum conversion and inbred development on genome diversity as revealed by high-resolution genotyping. Crop Sci 48:S-12–S-26
Koenning S, Overstreet C, Noling J, Donald P, Becker J, Fortnum B (1999) Survey of crop losses in response to phytoparasitic nematodes in the United States for 1994. J Nematol 31:587–618
Kollo I (2002) Plant-parasitic nematodes of sorghum and pearl millet: Emphasis on Africa. In: Leslie J (ed) Sorghum and Millets Diseases. Iowa State Press, pp 259–266
Krattinger SG, Lagudah ES, Spielmeyer W, Singh RP, Huerta-Espino J, McFadden H, Bossolini E, Selter LL, Keller B (2009) A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. Science 323:1360–1363
Li XH, Wang ZH, Gao SR, Shi HL, Zhang SH, George MLC, Li MS, Xie CX (2008) Analysis of QTL for resistance to head smut (*Sporisorium reiliana*) in maize. Field Crops Res 106:148–155
Lin YR, Schertz KF, Paterson AH (1995) Comparative analysis of QTLs affecting plant height and maturity across the Poaceae, in reference to an interspecific sorghum population. Genetics 141:391–411
Little C, Magill C (2009) The grain mold pathogen, *Fusarium thapsinum*, reduces caryopsis formation in *Sorghum bicolor*. J Phytopathol 157:518–519
MacClure M, Husman S, Schmitt M (1998) Infection of sorghum varieties by the cotton root-knot nematode, *Meloidogyne incognita*. Cotton, A College of Agriculture Report. University of Arizona, Tucson, In
Marcy DE (1937) Inheritance of resistance to the loose and covered kernel smuts of sorghum II. Feterita hybrids. Bull Torrey Bot Club 64:245–267

McSorley R, Wang K-H, Dover K (2004) Nematode management using sorghum and its relatives, ENY 716. University of Florida, Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences, Gainesville, FL

Mehta PJ, Wiltse CC, Rooney WL, Collins SD, Frederiksen RA, Hess DE, Chisi M, TeBeest DO (2005) Classification and inheritance of genetic resistance to anthracnose in sorghum. Field Crops Res 93:1–9

Melchers L, Lowe A (1943) The development of sorghums resistant to milo disease. University KS Technical Bulletin 55. Kansas State College Press, Manhattan, KS

Menz MA, Klein RR, Mullet JE, Obert JA, Unruh NC, Klein PE (2002) A high-density genetic map of *Sorghum bicolor* (L.) Moench based on 2926 AFLP, RFLP and SSR markers. Plant Mol Biol 48:483–499

Mittal M, Boora K (2005) Molecular tagging of gene conferring leaf blight resistance using microsatellites in *Sorghum bicolor* (L.) Moench]. Indian J Exp Biol 43:262–265

Muriithi LM, Claflin LE (1997) Genetic variation of grain sorghum germplasm for resistance to *Pseudomonas andropogonis*. Euphytica 98:129–132

Nafziger T (1918) How sorghum crosses are made. J Hered 9:321–322

Naidoo G, RA F, Torrez-Montalvo J, Magill C (1999) Maize and sorghum isolates of *Sporoisorium reilianum* differ in electrophoretic karyotype. MMPOL http://www.bspp.org.uk/mppol/1999/0419naidoo/. Accessed 03 June 2012

National Research Council (1996) Lost crops of Africa. National Academy Press, Washington DC

Nicholson R, Kollipara S, Vincent J, Lyons P, Cadena-Gomez G (1987) Phytoalexin synthesis by the sorghum mesocotyl in response to infection by pathogenic and nonpathogenic fungi. Proc Natl Acad Sci USA 84:5520–5524

Njugama J (2001) Combating heead smut of maize caused by *Sphacelotheca reiliana* through resistance breeding. In: Palmer A (ed) Seventh Eastern and Southern Africa regional maize conference. CIMMYT, Nairobi, Kenya, pp 110–112

Oh BJ, Frederiksen R, Magill C (1996) Identification of RFLP markers linked to a gene for downy mildew resistance (*sdm*) in sorghum. Can J Bot 74:315–317

Parh D, Jordan D, Aitken E, Mace E, Jun-ai P, McIntyre C, Godwin I (2008) QTL analysis of ergot resistance in sorghum. Theoret Appl Genet 117:369–382

Pascual CB, Raymundo AD, Hyakumachi M (2000) Resistance of sorghum line CS 621 to *Rhizoctonia solani* AG1-ia and other sorghum pathogens. J Gen Plant Path 66:23–29

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboobur R, Ware D, Westhoff P, Mayer KFX, Messing J, Rokhsar DS (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556

Peterson GC, Schaefer K, Pendleton BB (2009) Registration of 16 sorghum germplasm lines. J Plant Regist 3:203–205

Prom LK, Erpelding JE, Montes-Garcia N (2007) Chinese sorghum germplasm evaluated for resistance to downy mildew and anthracnose. Comm Biom Crop Sci 2:26–31

Prom LK, Erpelding JE, Montes-Garcia N (2008) Evaluation of sorghum germplasm from China against *Claviceps africana*, causal agent of sorghum ergot. Plant Health Progress. http://naldc.nal.usda.gov/download/21185/PDF. doi:10.1094/PHP-2008-0519-01-RS

Prom LK, Montes-Garcia N, Erpelding JE, Perumal R, Medina-Ocegueda S (2010) Response of sorgum accessions from Chad and Uganda ro natural infection by the downy mildew pathogen, *Peronosclerospora sorghi* in Mexico and the USA. J Plant Dis Prot 117:2–8

Prom L, Perumal R, Erattainmuthu S, Erpelding J, Montes N, Odvody G, Greenwald C, Jin Z, Frederiksen R, Magill C (2011) Virulence and molecular genotyping studies of *Sporisorium reilianum* isolates in sorghum. Plant Dis 95:523–529

Ramasamy P, Menz M, Mehta P, Katilé S, Gutierrez-Rojas L, Klein R, Klein P, Prom L, Schlueter J, Rooney W, Magill C (2009) Molecular mapping of *cg1*, a gene for resistance to anthracnose (*Colletotrichum sublineolum*) in sorghum. Euphytica 165:597–606

Reyes L, Frederiksen R, Thakur H (1969) Anthracnose incidence on grain sorghum in the south Texas costal bend area in 1968. In: Proceedings of the sixth biennial grain sorghum research and utilization conference. Grain Sorghum Producers Association, Lubbock, TX, Amarillo, TX, pp 8–9

Rich PJ, Grenier C, Ejeta G (2004) Striga resistance in the wild relatives of sorghum. Crop Sci 44:2221–2229

Sangduen N, Hanna WW (1984) Chromosome and fertility studies on reciprocal crosses between two species of autotetraploid sorghum: *Sorghum bicolor* (L.) Moench and *S. halepense* (L.) Pers. J Hered 75:293–296

Schertz KF, Ritchey JM (1978) Cytoplasmic-genic male-sterility systems in sorghum. Crop Sci 18:890–893

Sindhu A, Maier T, Mitchum R, Hussey R, Davis E, Baum T (2009) Effective and specific *in planta* RNAi in cyst nematodes: Expression interference of four parasitism genes reduces parasitic success. J Exp Bot 60:315–324

Singh M, Chaudhary K, Singal H, Magill C, Boora K (2006) Identification and characterization of RAPD and SCAR markers linked to anthracnose resistance gene in sorghum [*Sorghum bicolor* (L.) Moench]. Euphytica 149:179–187

Sontheimer E (2005) Assembly and function of RNA silencing complexes. Na Rev Mol Cell Biol 6:127–138

Sriwatanapongse S, Junahyon S, Vasal S (1993) Suwan-1 maize from Thailand to the world. CIMMYT, Mexico, D.F

Stephens JC, Holland RF (1954) Cytoplasmic male-sterility for hybrid sorghum seed production. Agron J 46:20–23

Stephens JC, Miller FR, Rosenow DT (1967) Conversion of alien sorghums to early combine genotypes. Crop Sci 7:396

Swanson AF, Parker JH (1931) Inheritance of smut resistance and juiciness of stalk in the sorghum cross, Red Amber × Feterita. J Hered 22:51–56

Takken FLW, Tameling WIL (2009) To nibble at plant resistance proteins. Science 324:744–746

Tao YZ, Jordan DR, Henzell RG, McIntyre CL (1998) Identification of genomic regions for rust resistance in sorghum. Euphytica 103:287–292

Tesso TT, Claflin LE, Tuinstra MR (2005) Analysis of stalk rot resistance and genetic diversity among drought tolerant sorghum genotypes. Crop Sci 45:645–652

Toler R (1985) Maize dwarf mosaic virus, the most important virus disease of sorghum. Plant Dis 69:1011–1015

Tuinstra MR, Soumana S, Al-Khatib K, Kapran I, Toure A, van Ast A, Bastiaans L, Ochanda NW, Salami I, Kayentao M, Dembele S (2009) Efficacy of herbicide seed treatments for controlling Striga infestation of sorghum. Crop Sci 49:923–929

Vallad GE, Goodman RM (2004) Systemic acquired resistance and induced systemic resistance in conventional agriculture. Crop Sci 44:1920–1934

Wrather A, Sweets L (2009) Management of grain sorghum diseases in Missouri. Missouri University Extension. http://extension.missouri.edu/publications/DisplayPub.aspx?P=G4356. Accessed 03 June 2012

Xu G, Cui Y, Schertz K, Hart G (1995) Isolation of mitochondrial DNA sequences that distinguish male-sterility-inducing cytoplasms in *Sorghum bicolor* (L.) Moench. Theor Appl Genet 90:1180–1187

Yadav BC, Veluthambi K, Subramaniam K (2006) Host-generated double stranded RNA induces RNAi in plant-parasitic nematodes and protects the host from infection. Mol Biochem Parasitol 148:219–222

Yonemaru J-i, Ando T, Mizubayashi T, Kasuga S, Matsumoto T, Yano M (2009) Development of genome-wide simple sequence repeat markers using whole-genome shotgun sequences of sorghum (*Sorghum bicolor* (L.) Moench). DNA Res 16:187–193

Zhao B, Lin X, Poland J, Trick H, Leach J, Hulbert S (2005) A maize resistance gene functions against bacterial streak disease in rice. Proc Natl Acad Sci USA 102:15383–15388

Zipfel C, Kunze G, Chinchilla D, Caniard A, Jones JDG, Boller T, Felix G (2006) Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts Agrobacterium-mediated transformation. Cell 125:749–760

Chapter 16
Bridging Conventional and Molecular Genetics of Sorghum Insect Resistance

Yinghua Huang, Hari C. Sharma, and Mukesh K. Dhillon

Abstract Sustainable production of sorghum, *Sorghum bicolor* (L.) Moench, depends on effective control of insect pests as they continue to compete with humans for the sorghum crop. Insect pests are a major constraint in sorghum production, and nearly 150 insect species are serious pests of this crop worldwide and cause more than 9% loss annually. Annual losses due to insect pests in sorghum have been estimated to be $1,089 million in the semiarid tropics (ICRISAT Annual report 1991. International Crop Research Institute for Semi-arid Tropics. Patancheru, Andhra Pradesh, India, 1992), but differing in magnitude on a regional basis. Key insect pests in the USA include the greenbug, *Schizaphis graminum* (Rondani); sorghum midge, *Stenodiplosis sorghicola* (Coquillett); and various caterpillars in the Southern areas. For example, damage by greenbug to sorghum is estimated to cost US producers $248 million annually. The major insect pests of sorghum on a global basis are the greenbug, sorghum midge, sorghum shoot fly (*Atherigona soccata* Rond.), stem borers (*Chilo partellus* Swin. and *Busseola fusca* Fuller), and armyworms (*Mythimna separata* Walk and *Spodoptera frugiperda* J.E. Smith). Recent advances

---

Mention of a trademark or proprietary product does not constitute a guarantee or warranty of a product by the US Department of Agriculture and does not imply its approval to the exclusion of other products that may also be suitable.

Y. Huang (✉)
USDA-ARS Plant Science Research Laboratory,
1301 N. Western Road, Stillwater, OK 74075, USA
e-mail: yinghua.huang@ars.usda.gov

H.C. Sharma
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT),
Patancheru 502324, Andhra Pradesh, India
e-mail: h.sharma@cgiar.org

M.K. Dhillon
Division of Entomology, Indian Agricultural Research Institute (IARI),
New Delhi 110012, India
e-mail: mukeshdhillon@iari.res.in

A.H. Paterson (ed.), *Genomics of the Saccharinae*, Plant Genetics and Genomics: Crops and Models 11, DOI 10.1007/978-1-4419-5947-8_16,

in sorghum genetics, genomics, and breeding have led to development of some cutting-edge molecular technologies that are complementary to genetic improvement of this crop for insect pest management. Genome sequencing and genome mapping have accelerated the pace of gene discovery in sorghum. Other genomic technologies, such as QTL (quantitative trait loci) mapping, gene expression profiling, functional genomics, and gene transfer are powerful tools for efficient identification of novel insect-resistance genes, and characterization of the key pathways that regulate the interactions between crop plants and insect pests leading to successful expression of the host plant defense. Traditional breeding methods, such as germplasm evaluation and enhancement, backcrossing, pedigree selection, and recurrent selection continue to play important roles in developing insect-resistant cultivars with major resistance genes; and new cultivars with enhanced resistance to several important insect pests are released continuously. Future research efforts should focus on identification of new sources of resistance, characterization of resistance genes, and dissecting the network of resistance gene regulation. Collaboration between research institutions and the sorghum industry as well as international cooperation in utilization of emerging knowledge and technologies will enhance the global efforts in insect pest management in sorghum.

1 Introduction

Sorghum [*Sorghum bicolor* (L.) Moench] is grown in warm and humid environments that often favor the proliferation of many insect pests. In addition, long growth periods of some varieties such as sweet sorghum and forage sorghum, cultivation of genetically homogeneous hybrids/varieties on large acreages, and the practice of multicropping sorghum and rotation by intercropping with common host plants throughout the years result in the buildup of pest populations. Thus, insect pest populations often exceed the economic-injury level in most of the sorghum-growing areas.

2 Insect Pests of Sorghum

Numerous insects attack sorghum worldwide. More than 150 insect species are considered pests of sorghum (Sharma 1993), and about 20 of them can cause severe economic damage to the crop. A summary of the major insect pests of sorghum is listed in Table 16.1 and described in detail, including their geographical distribution, biology, feeding characteristics, and symptoms of damage. Like most field crops, sorghum is usually attacked by only a few key pests in each agro-ecosystem (Young and Teetes 1977). Primary insect species on sorghum vary depending upon the location and the environments. Depending on sorghum agro-ecosystems, one or a few key insect pests, such as greenbug, *Schizaphis graminum* (Rondani), sorghum

Table 16.1 Major insect pests of sorghum

Common name	Scientific name	Geographical distribution	Feeding characteristics and damage
African (Nutgrass) Armyworm	*Spodopera exempta*	East and West Africa	Damage caused by larvae, gregarious, move through at high rates.
Banks Grass Mite	*Oligonychus pratensis*	USA, India	Suck sap from underside of leaves, infestation spreads upwards, web panicles, lodging.
Chinch Bug	*Blissus leucopterus*	North America	Withdraw enormous quantities of sap from stems, reddened, weak, stunted, lodge.
Corn Earworm	*Helicoverpa zea*	Americas	Feed on tender, folded leaves and developing grain.
Corn Leaf Aphid	*Rhopalsiphum maidis*	Sorghum-growing areas of world	Suck plant juice, yellowish mottling of leaves, honeydew production hinders harvesting.
Earhead (Christmas Berry) Webworm	*Cryptoblabes gnidiella*	India	Destroy grain, webs of silken thread on inside and outside of head.
Earhead Bug	*Calocoris angustatus*	India, Africa	Infest panicles, suck sap from developing grain, shriveled and punctured grain.
Earhead Webworm	*Nola analis*	Africa, Asia	Larva feeds on grain in the head.
Fall Armyworm	*Spodoptera frugiperda*	Southeastern USA, tropical America	Feed on tender parts of whorl leaves and developing grain of panicle after emergence.
Grasshoppers	*Oedaleus senegalensis*, *Aliopus simulatrix*	Africa (southern Sahara borders), arid parts of world	Attacked at all plant stages, most loss at seedling stage or ripening of panicles.
Greenbug	*Schizaphis graminum*	Asia, Africa, Australia, America	Injects toxins, reddish spots on leaves, seedling pest, damaging at heading stage.
Maize Stalk Borer	*Busseola fusca*	Africa	Deadhearts, growth retarded, reduced grain and flowering.
Oriental Armyworm	*Mythimna separata*	Asia, Pacific Islands, Austrialia, Fiji, New Zealand	Feed on leaves leaving only midrib, immature panicle damage, entire crop loss if heavy infestation.
Shoot Bug	*Peregrinus maidis*	India, Africa, West Indies, Bermuda, Philippines	Sap-sucking, stunted growth, leaf death and sometimes whole plant, prevent panicle emergence.
Shoot Fly	*Atherigona soccata*	Semiarid tropics, but not in Americas and Australia	Larval feeding on central leaf, deadheart symptom, removed deadheart emits a bad smell.

(continued)

Table 16.1 (continued)

Common name	Scientific name	Geographical distribution	Feeding characteristics and damage
Sorghum Midge	*Contarinia sorghicola*	Worldwide except Southeast Asia	Larvae feeding on the ovary, "blasted panicle."
Sorghum Webworm	*Nola sorghiella*	Humid south of USA and Central America	Feed on developing floral parts, gnaw circular holes in the seed.
Spotted Stem Borer	*Chilo partellus*	East Africa, Indian subcontinent, and the Far East	Small elongated windows in young whorl leaves, bore into the stem, deadheart symptoms, chaffy seeds.
Sugarcane Aphid	*Melanaphis sacchari*	Asia, Africa, tropical America	Prefers older leaves, sap-sucking causing stunted plant growth, drying of leaves and death.
Sugarcane Borer	*Diatraea saccharalis, D. lineolata, D. grandiosella*	North and South America	Bore up and down stalk, lodging, panicle breakage.
Sugarcane Borer	*Eldana saccharina*	Africa south of the Sahara	Larvae hang by silken threads, feed on leaves, bore midrib, deadhearts.
Yellow Sugarcane Aphid	*Sipha flava*	New World	Prefers older lower leaves, secretes potent toxin killing seedlings, purpling/yellowing leaves, stunting, lodging.

midge, *Stenodiplosis sorghicola* (Coquillett), shoot fly, *Atherigona soccata* (Rondani), or stem borer, *Chilo partellus* (Swinhoe), may occur perennially and dominate the pest control practices. In addition to the key pests, there are numerous insect species that can cause periodic plant damage and/or yield loss. Many other insects reported to damage sorghum are of regional/local importance or are only occasional pests that cause economic damage only in localized areas or only during some years.

Insect pests usually attack certain parts of sorghum plants (Teetes and Pendleton 2000). Insects attacking planted seeds and roots include wireworms, red fire ants, and white grubs. Insect pests of sorghum seedlings are cutworms, southern corn rootworm, chinch bug, and shoot fly. Insect pests often feeding on leaves and leaf whorls are greenbug, corn leaf aphid, yellow sugarcane aphid, shoot bug, oriental armyworm, whorl-infesting caterpillars, grasshoppers, banks grass mite, and fall armyworm. Sorghum panicle-feeding insects include sorghum midge, corn earworm, earhead worm, fall armyworm, sorghum webworm, and African head bug. Stalk-damaging insects are sugarcane borer, spotted stem borer, southwestern corn borer, Mexican rice borer, sugarcane rootstock weevil, pink borer, and African sugarcane borer. Insect pests of sorghum grain include grain weevil, grain borer, grain

moth, grain beetles, flour beetles, and flour moth. Some insect pests feed on several plant parts. For example, fall armyworm larvae cause extensive defoliation of the crop and also burrow into the growing point (bud, whorl, etc.), destroying the growth potential of plants. Fall armyworms infest panicles after panicle emergence, in the same manner as corn earworm, *Helicoverpa* zea (Boddie). Some insect species also transmit viral diseases.

In spite of the impressive gains in sorghum production over the years, biotic and abiotic stresses continue to haunt sorghum farmers across the globe. Grain yields in the semiarid tropics (SAT) are generally low (500–800 kg ha^{-1}) mainly due to insect pest damage (Sharma 1993). Accurate assessment of sorghum grain yield losses due to insect attack is scarce and difficult to obtain, however, annual losses to insect pests in sorghum have been estimated to be $1,089 million in the semiarid tropics (ICRISAT 1992). In India, nearly 32.1% of the actual produce is lost due to insect pests (Borad and Mittal 1983), although they differ in magnitude on a regional basis.

As mentioned above, there are four devastating insect pests on sorghum, which cause severe damages to the crop and serious economic losses to sorghum producers. Sorghum midge, *S. sorghicola*, is one of the most ubiquitous and damaging insect pests attacking sorghum. It has been the subject of research since its first discovery in 1894 in Queensland (Tryon 1895). Estimates of annual cost to producers are $28 million in Texas (Peterson et al. 1997), $294 million in the semiarid tropics (ICRISAT 1992), and $10 million in Australia (Henzell and Jordan 2009). A breeding program for resistance to sorghum midge began at Texas A&M University soon after usable resistance sources of resistance were identified (Johnson et al. 1973), and similar research programs began at the Queensland Department of Primary Industries, Australia in 1975 (Henzell et al. 1980), and at ICRISAT Asia Center in India in 1980 (Sharma et al. 1994). In addition, the private sectors in Australia and the USA have also been breeding for resistance to sorghum midge.

The greenbug, *S. graminum* has been a major pest of sorghum since 1968 when biotype C was first reported (Harvey and Hackerott, 1969). Small grains, primarily wheat, act as winter hosts, and where the growing season of small grains does not overlap with that of sorghum, grasses such as Johnson grass, *Sorghum halepense*, serve as interim hosts. Sorghum growers have supported research on development of greenbug-resistant sorghums, and have benefitted from the resistant cultivars produced. The US and foreign consumers experienced economic benefits amounting to $248 million and $274 million annually, respectively (Eddleman et al. 1999).

Shoot fly, *A. soccata* is an important pest of sorghum in Asia, Mediterranean Europe, and Africa. Annual losses have been estimated at $200 million. In India, shoot fly damage in sorghum at times results on 90% reduction in grain yield, and 45% of fodder yield, (ICRISAT 1992). The severity of damage by spotted stem borer, *C. partellus* can result in severe loss of crop stand when seedlings are attacked. It is a serious pest in Asia, and East and southern Africa. Maize stalk borer, *B. fusca*, is a major pest in the African highlands. The stem borers can also infest the plants at a later stage, causing stem tunneling, which weakens the stem and results in stem

breakage and unfilled grains. Losses caused by stem borers have been reported to be between 5 and 15% in West Africa, and 18 and 27% in East Africa. In India, reported losses range as high as 55–83% on certain susceptible hybrids and varieties during severe infestations (ICRISAT 1989). In the semiarid tropics, stem borers cause an annual loss of $334 million (ICRISAT 1992).

3 Sorghum Plant Response to Insect Attack

There are many harmful insects that coexist with plants in various agro-ecosystems. During their coevolution with plants, insects have evolved the ability to search their host plants for feeding and oviposition using physical and/or chemical cues from the host plants. During the long course of interactions, herbivorous insects developed a compatible relationship with their plant hosts, so they are able to live on those plants. Based on their interactions with the host plants, there are two types of insects: generalists and specialists. Generalist herbivorous insects have a wide host range, being able to feed on many species of plants; whereas specialists have a narrow host range, attacking only one or a few plant species within the same family. Greenbug and stem borers are good examples of generalist insect pests, and shoot fly and midge as specialists on sorghum. Plant injury occurs when insect feeding causes abnormal metabolism and function, leading to irreversible physical or chemical changes in plant. In general, plant injury is expressed as defoliation, reduction in growth, yield loss, or poor grain quality.

On the contrary, plants have also developed their defense systems (i.e., host plant resistance [HPR]) to counteract herbivore attack. Defense is costly; thus discerning insect feeding from causal mechanical wounding and quickly deploying increased levels of defensive compounds that are critical to effectively battle insect pests infesting sorghum plants (Huang 2007; Park et al. 2006). As a result, insect-resistant plants can alter the relationship that an insect pest has with its plant host. The interactions between the insects and plants are dependent upon different resistance mechanisms. Three types (i.e., mechanisms) of resistance were described by Painter (1951), including antibiosis, antixenosis (non-preference), and tolerance. *Antibiosis* affects the biology of the insect so that pest abundance and subsequent damage is reduced as compared to the one that would have occurred normally if the insect feeds on a susceptible crop variety. Antibiosis resistance often results in increased mortality or prolonged development and rescued fecundity. *Antixenosis* affects the behavior of an insect pest and usually is expressed as non-preference of the insect for the resistant plant as compared to a susceptible plant. *Tolerance* is the capability of a host plant to withstand or recover from the damage caused by insect pest abundance that would damage a susceptible plant. Tolerance is a plant response to an insect pest and differs from antibiosis and antixenosis resistance as to how it affects the insect–plant relationship. Antibiosis and antixenosis components of resistance induce an insect response when the insect attempts to use the resistant plant for food, oviposition, or shelter. Antibiosis, antixenosis, and tolerance are the major

components of resistance against sorghum stem borer, *C. partellus* and maize stalk borer, *B. fusca* (van den Berg et al. 1994; Sharma and Nwanze 1997). Greenbug resistance in sorghum has often been explained as tolerance, but our recent studies with diverse resistance sorghum germplasm suggested that both antibiosis and antixenosis are present in some resistant lines of sorghum (Huang, unpublished data).

Host plant defenses can also be described as constitutive or inducible. Constitutive defenses include physical and chemical barriers or traits that are formed regardless of the presence of insects. For instance, some plants form many external structural barriers such as sharp prickles, thorns, trichomes, and cuticles that restrict insect attack and feeding. Other plants produce and release defense compounds such as resins, lignins, and wax that alter the texture of the plant tissues which discourage herbivory. Some sorghum varieties are able to synthesize cyanogenic glycosides and store them in inactive forms in plant vacuoles. They become toxic when herbivores eat the plant and break the cell membranes, and thus allowing the glycosides to come into contact with enzymes in the cytoplasm that catalyze reactions releasing hydrogen cyanide, which blocks cellular respiration of the insects. Phenolics are also shown to have negative effects on herbivores, while the condensed tannins—polymers composed of 2–50 (or more) flavonoid molecules, can inhibit herbivore digestion by binding to consumed plant proteins. Constitutive and inducible defenses are achieved through similar means, but differ in the sense that constitutive defenses are expressed before insect attack, while inducible defense is activated only after a plant is attacked by an insect.

HPR has often been used for successful management of several insect pests in sorghum, but certain limitations and problems will always beset any insect control program, and HPR is no exception. Although several sources of resistance have been identified against greenbug, shoot fly, spotted stem borer, sorghum midge, and head bugs in sorghum (Sharma et al. 1988a, b, 1992, 2003; Kumari et al. 2000; Huang 2004, 2011), only a few of them are being deployed in the development of insect-resistant varieties, as it takes a long time and needs a great deal of expertise and resources. In a complementary approach, efforts have been made in the past toward the development of insect-resistant sorghums using biotechnological approaches, discussed in a later section. Good examples of the recent advances in development of insect-resistant sorghums include the effects of cytoplasmic male-sterility on expression of resistance to insect pests, transfer of insect-resistance genes in cytoplasmic male-sterile (CMS/A), maintainer (B), and restorer (R) lines, diversification of CMS systems, and development of insect-resistant transgenic sorghums.

4 Identification of Resistance Sources and Utilization of HPR

HPR is a pest management method that utilizes the plant's own defense mechanisms against an insect pest. The prerequisite for this approach is to identify sources of resistance. As sorghums are native to the warm environments, where

their associated herbivores and entomophages have coevolved, there is an opportunity to search for insect-resistance sources in the diverse sorghum germplasm. Sorghum genetic resources are conserved at many centers around the world as detailed by Kimber (2012).

Accessions of these sorghum seed collections were obtained from all geographic regions in the world, from tropical to temperate zones, from high and low elevations, as well as from those grown in different seasons of the year. Thus, any resistance sources can probably be found in such world sorghum germplasm collections, although much of them have not yet been evaluated for resistance to insects. Greenbug is a key pest of sorghum in most areas of the USA where the crop is grown, especially in the Great Plains. To manage this damaging pest, sources of resistance were found in sorghum germplasm in the late 1960s (Young and Teetes 1977), and since then efforts in finding new sources continue at many public research laboratories (Peterson et al. 1997; Huang 2006). Releases of resistant breeding materials have been made to commercial seed companies. Thus, incorporation of greenbug resistance into elite parental lines and hybrids has been the goal of every sorghum breeding program in the USA. However, the effectiveness of these resistant germplasms has been limited by the tremendous diversity of virulence genes (i.e., biotypes) that exists within greenbug populations (Burd and Porter 2006). The original biotype to attack sorghum was biotype C, and since then, 10 additional biotypes have been detected, and three of these (E, I, and K) damage sorghum (Peterson et al. 1997). Several examples of successful deployment of resistant cultivars can be cited, but resistance is oftentimes short-lived due to the tremendous diversity of biotypes in the target pest populations and evolution of virulence. New sources of resistance to these key pests must be found continuously and incorporated into high-performance breeding lines for cultivar/hybrid development. Continuous improvement in crop defense against the new biotypes of the greenbug is dependent on the availability of diverse genetic resources and judicious use of effective sources of resistance. At present, over 40,000 sorghum germplasm accessions, including many exotic sources, have been evaluated for their response to greenbug feeding in the greenhouse at the USDA-ARS Plant Science Research Laboratory, Stillwater, Oklahoma, leading to the identification of new sources of resistance and novel genetic resistance in sorghum (Huang 2011).

In India, over the past five decades, a large proportion of the world sorghum germplasm collection has been evaluated for resistance to insect pests, and a number of lines with resistance to the major insect pests have been identified (Sharma et al. 1992, 2003). Large-scale screening of the sorghum germplasm at ICRISAT has resulted in identification of several lines with reasonable levels of resistance to shoot fly, stem borer, midge, and head bugs. Sources of resistance to insects in sorghum have been used in the breeding program, and many varieties with resistance to insect pests have been developed. However, cultivars with resistance to insect pests are cultivated by the farmers only on a limited scale due to over emphasis on grain yield as a criterion to release cultivars by the national programs. Now having achieved a plateau in grain yield in sorghum, it is important that insect resistance be used as one of the criteria to identify varieties and hybrids for use by the farmers for

sustainable crop production. To effect this change, the research needs to be driven more by the ground reality than by perception.

Identification and transfer of insect resistance from the wild relatives of sorghum have been much less successful than that of resistance to diseases. Nonetheless, the wild relatives of sorghum provide sources of diverse genes for resistance to insect pests (Sharma et al. 2005). The vast and largely underutilized pool of desired genes/traits existing in the wild relatives of cultivated sorghum will provide a huge new resource of genetic resistance to promote the next phase of sorghum genetic improvement for insect resistance.

Levels of resistance to sorghum shoot fly (*A. soccata*) and stem borer (*C. partellus*) in the cultivated sorghum are low to moderate (Sharma et al. 1992, 2003). Therefore, it may be important to identify wild relatives of sorghum with high levels of resistance to insect pests (Venkateswaran et al. 2009). Wild species of sorghum (*Sorghum purpeosericeum* and *S. versicolor*) possess very high levels of resistance to shoot fly (Mote 1984). Venkateswaran (2003) identified several species of sorghum with high levels of resistance to *A. soccata*, with resistance levels close to immunity under field conditions. Accessions belonging to Parasorghum (*S. australiense*, *S. purpureosericeum*, *S. brevicallosum*, *S. timorense*, *S. versicolor*, *S. matarankense*, and *S. nitidum*) and Stiposorghum (*S. angustum*, *S. ecarinatum*, *S. extans*, *S. intrans*, *S. interjectum*, and *S. stipoideum*) did not show any shoot fly damage under multi-choice conditions in the field (Venkateswaran et al. 2009). Heterosorghum (*S. laxiflorum*) and Chaetosorghum (*S. macrospermum*) showed very low damage by the sorghum shoot fly. Within section Sorghum, the four wild races belonging to *S. bicolor* subsp. *verticilliflorum* (*aethiopicum*, *arundinaceum*, *verticilliflorum*, and *virgatum*) were highly susceptible to shoot fly, as was *S. halepense*. Fifteen species of wild relatives of sorghum have shown high levels of resistance to spotted stem borer, *C. partellus*, under artificial infestation in the field (Venkateswaran 2003). Species belonging to Heterosorghum (*S. laxiflorum*), Parasorghum (*S. australiense*, *S. purpureosericeum*, *S. versicolor*, *S. matarankense*, *S. timorense*, *S. brevicallosum*, and *S. nitidum*), and Stiposorghum (*S. angustum*, *S. ecarinatum*, *S. extans*, *S. intrans*, *S. interjectum*, and *S. stipoideum*) showed little damage by the spotted stem borer larvae, except for one accession of Heterosorghum, which showed 2% deadhearts. In contrast, section Chaetosorghum (*S. macrospermum*) was highly susceptible to stem borer damage. Within section Sorghum, the four wild races of *S. bicolor* subsp. verticilliflorum (races arundinaceum, aethiopicum, verticilliflorum, and virgatum) were highly susceptible to stem borer damage, as was *S. halepense*. Sorghum midge, *S. sorghicola* females did not lay any eggs in the spikelets of wild relatives of sorghum such as *S. angustum*, *S. amplum*, and *S. bulbosum* compared to 30 eggs in *S. halepense* under no-choice conditions (Sharma and Franzmann 2001). Larger numbers of sorghum midge females were attracted to the odors from the panicles of *S. halepense* than to the odors from panicles of *S. stipoideum*, *S. brachypodum*, *S. angustum*, *S. macropsermum*, *S. nitidum*, *S. laxiflorum*, and *S. amplum*. The accessions belonging to the secondary gene pool with diverse mechanisms of resistance can be crossed with cultivated sorghum, while those belonging to the tertiary gene pool may require

application of embryo rescue techniques to transfer resistance genes from the wild relatives into cultivated sorghums.

Use of insect-resistant crop varieties is economically, ecologically, and environmentally advantageous. Economic benefits occur because crop yields are saved from loss due to insect pests and money is saved by not applying insecticides that would have been applied to susceptible varieties. In most cases, seed of insect-resistant cultivars costs no more, or a little more, than for susceptible cultivars. Ecological and environmental benefits would certainly arise from increases in species diversity in the agro-ecosystem, in part because of reduced use of insecticides.

5 Conventional Approaches of Insect Pest Management

The primary goals of conventional breeding of sorghum have been focused on the improvement of the crop with higher yield and better quality. Once this is achieved, resistance traits can be incorporated, provided the methods for introducing the resistance can be readily integrated into the breeding programs. Human involvement in the improvement of sorghum began with identification and selection of plants with desirable characteristics for a better production. These new genotypes arose from random outcrosses or mutations that were fixed due to the self-pollination of the new type. Thus, a number of sorghum cultivars and elite breeding lines characterized for resistance to major insects have been developed through conventional breeding methods (Rooney 2004). In such breeding approaches, particularly in breeding insect resistance, sorghum breeders usually search for genetic variability for insect resistance and then incorporate the desired traits into breeding lines, leading to the development of resistant commercial cultivars or hybrids. Since sorghum is a self-pollinated species, most breeding methodologies (both cultivar and hybrid) are based on the production of segregating populations followed by selection in segregating populations. The selections are usually allowed to self-pollinate during selection to produce homozygous uniform lines (i.e., pure-line cultivars). In hybrid breeding programs, these lines will be test crossed to measure their value as a parental line. Later, the use of hybrid vigor is facilitated by cytoplasmic male-sterility. Evidently, incorporation of insect resistance into elite parental lines and hybrids has been the goal of many sorghum breeding programs in the USA and elsewhere. The goal of most population improvement programs is to accumulate favorable alleles for the traits of interest while maintaining as much genetic diversity as possible. Wide hybridization involving crosses between cultivated sorghum and un-adapted sorghum germplasm lines or related wild species is also used in some breeding programs. The benefits of wide hybridization include broadening the genetic diversity of the crop and utilizing the newly identified resistance genes in distantly related sources.

Insect pests of sorghum should be managed by actions that prevent insects from becoming abundant enough to cause economic damage. Management actions include using nonchemical and chemical methods. Cultural and biological

management methods are nonchemical methods imposed to avoid insect pests, suppress insect pest abundance or rate of increase, delay the time when insect pests reach damaging abundance levels, or increase plant tolerance to insect pests. Because nonchemical management methods are imposed to avoid or prevent insect pest damage, the decision to use these methods must be made before an insect pest problem occurs, and often even before the crop is planted. Chemical management methods involve use of insecticides to kill insect pests. Insecticides have a rapid and curative action but are costly and may cause negative ecological and environmental consequences. Therefore, their use must be justified based on actual measurement of insect abundance and damage. Integrated pest management (IPM) involves the use of a combination of management methods in a strategy to maintain insect pest abundance or damage below levels that cause economic loss. IPM has been a practical approach to dealing with insect pests of sorghum.

To manage insect pests of sorghum, integrated pest management systems depend on an array of different approaches, such as planting time to avoid infestation, crop rotation, biological control by using natural enemies, chemical control, etc. In cases where cultural control methods are not always effective and pesticides are either not available or economically feasible, the development and deployment of genetically resistant crops in an integrated pest management production system is an effective, economical, and environmentally sound approach (Sharma 1993). In addition, breeding for crop resistance to insect pests was considered to be a safe and inexpensive insect control method, highly applicable to small-farm conditions and easily accepted by sorghum producers. In recent years, research has focused on identification of insect-resistant sources and various sorghum improvement programs have intensified their efforts to breed insect-resistant and stable-yielding varieties and hybrids as these are likely to be more critical for sustainable production.

Cytoplasmic male-sterility has an effect on expression of resistance to insect pests. Most of the sorghum hybrids grown across world are based on the A_1 cytoplasm, which has been found to be highly susceptible to insect pests (Sharma 2001; Sharma et al. 2004; Dhillon et al. 2008). Ross and Kofoid (1979) have also reported that the Kansas lines KS 34 to KS 39 based on Kansas CMS system are as susceptible as CKA (Combine Kafir-based CMS lines) to the greenbug, *S. graminum*. Sharma et al. (1994) recorded low damage by the sorghum midge, and reduced midge emergence on midge-resistant B-lines as compared to corresponding A-lines. However, no differences were found for midge damage or adult emergence between midge-resistant and -susceptible A-lines. Midge-resistant CMS x susceptible restorer-based hybrids were less susceptible to *S. sorghicola* damage than susceptible CMS × susceptible restorer-based hybrids (Johnson 1977; Sharma et al. 1996). The expression of non-preference and antibiosis components of resistance to southwestern corn borer, *Diatraea grandiosella* Dyar, and sugarcane borer, *Diatraea saccharalis* Fab., was higher in resistant inbred line-based hybrids, CML 67 × CML 135 and CML 139 × CML 135 than the inbreds (Kumar and Mihm 1996). These hybrids also suffered low leaf and stalk damage, and grain yield loss in comparison to the susceptible hybrid Ki 3 × CML 131. The oviposition and deadheart formation on main plants and tillers of sorghum by the shoot fly were significantly lower on

maintainer lines compared to the CMS lines (Dhillon et al. 2006a). Larval development was prolonged and pupal mortality was greater on maintainer lines than on the CMS lines, while pupal weights and fecundity were greater on the CMS lines (Dhillon et al. 2006a). The maintainer lines showed better recovery resistance than the CMS lines, but such differences were more apparent in the shoot fly-resistant CMS and maintainer lines than in the shoot fly-susceptible CMS and maintainer lines. Furthermore, the A_4M cytoplasm has been found to be comparatively less susceptible to *A. soccata* damage than the A_1, A_2, A_3, A_4G, or A_4VzM cytoplasms (Dhillon 2004; Dhillon et al. 2005). Expression of morphological traits such as leaf glossiness, trichomes, and leaf surface wetness was better in the maintainer lines as compared to the CMS lines (Dhillon et al. 2006c). The shoot bug- and sugarcane aphid-resistant CMS lines suffered more damage than the B-lines, while such differences were not apparent in case of susceptible CMS and maintainer lines (Dhillon et al. 2006b). The stem borer-resistant CMS and maintainer lines experienced similar levels of deadheart formation, while the stem borer-susceptible maintainers suffered more damage than the CMS lines (Dhillon et al. 2006b), indicating that the expression of resistance may be influenced by the interaction of factors in the cytoplasm of maintainer lines with the nuclear genes. Hybrids based on shoot bug, sugarcane aphid, midge, and shoot fly-resistant CMS and restorer lines suffered less damage than the hybrids based on susceptible CMS and resistant or susceptible restorer lines, suggesting that expression of resistance to these insects is influenced by the genetic background of the CMS lines (Sharma et al. 2004; Dhillon et al. 2006d; Sharma et al. 2006). However, the hybrids based on stem borer-resistant or -susceptible CMS lines with resistant restorers showed significantly lower deadheart formation as compared to the hybrids based on stem borer-resistant or susceptible CMS lines and -susceptible restorers, suggesting that restorer lines exercised a greater influence on expression of resistance to stem borer in sorghum (Dhillon et al. 2006b). Similar results have also been reported for expression of resistance to stem borers, *C. partellus* and *B. fusca* in maize (Kumar 1993). Dhillon et al. (2008) suggested that the genetic background of CMS, cytoplasmic factors, the interactions of the factors in the cytoplasm of maintainer lines with the nuclear genes and the restoration abilities of the restorers, influence the expression of resistance to insect pests in cereals depending on the crop and the insect species involved, and therefore, there is a continuing need to evaluate different cytoplasms for their effects on cultivar susceptibility to insect pests before being finally deployed in crop improvement programs. Also, there is an urgent need to convert various sources of resistance to insect pests into CMS, maintainer, and restorer lines, so as to be able to develop hybrids with increased levels and diverse mechanisms of resistance to target pests, which can be used by the public institutions and private seed industry to develop insect-resistant hybrids. Pedigree method is the most suitable and easy-to-operate breeding method for development of insect-resistant restorer and maintainer lines.

The commercial sorghum hybrids produced so far across the world are based on the single cytoplasm designated as A_1 (Reddy and Stenhouse 1994; Moran and Rooney 2003), except in China. However, based on experience in other crops (Tatum

1971), and the available information in sorghum, CMS lines are more susceptible to insect pests. Exclusive use of the A_1 cytoplasm as a source of male-sterility in commercial hybrid seed for multiplication of sorghum might restrict nuclear diversity of male-sterile (A) as well as restorer (R) lines, and thus, increase vulnerability to other biotic and abiotic stresses. Therefore, as a contingency plan to prevent such eventualities and to broaden the genetic base, several non-*milo* CMS systems designated as A_2, A_3, A_4 (VzM), A_4 (Maldandi), and A_4 (Guntur) were identified and developed (Schertz 1994) for use in hybrid breeding programs. Several nuclear genes are known to control expression of CMS and thus different CMS types can be distinguished through restoration patterns in testcrosses and anther morphology (classical method) or using molecular markers. Differentiation of cytoplasmic types, in addition to male-sterility influences, is a prerequisite for their efficient utilization to diversify the CMS-base of hybrids. Further, deployment of different CMS systems in different genetic backgrounds is necessary to diversify CMS systems in order to nullify the CMS effect and develop pest-resistant hybrids.

6 Genomic Tools for Genetic Analysis of Insect Resistance and Marker-Assisted Breeding

To facilitate the identification and isolation of chromosomal regions or genes involved in adaptation to adverse environments or resistance to biotic stresses, genome resources enabling analysis of sorghum trait loci have been developed. Genetic mapping of sorghum based on DNA markers began almost two decades ago and the first sorghum maps were based largely on DNA probes from maize (Hulbert et al. 1990). Although several versions of sorghum maps have been produced, they were not complete until the maps of Pereira et al. (1994) and Chittenden et al. (1994), which contain all ten linkage groups, known lately as ten chromosomes (Kim et al. 2005). Both of the later maps were constructed with RFLP markers in F_2 populations of relatively small size. After that, a more detailed RFLP-based map (Bowers et al. 2003), an AFLP-based map (Klein et al. 2000), and an integrated RFLP and SSR linkage map (Bhattramakki et al. 2000) were developed for sorghum.

Recently, simple sequence repeats (SSRs) or microsatellites have become the most important DNA marker technology as they proved to be a more dependable, rapid, and inexpensive tool for plant genotyping (Yang et al. 1996). We have recently constructed a detailed SSR-based genetic map for sorghum (Wu and Huang 2007), in which the mapped SSR loci distributed throughout all ten chromosomes and spanned a distance of 997.5 cM. With rapid increase in the availability of cDNA clones and expressed sequence tags (ESTs), we also took the in silico mining approach for the development of EST-SSRs (i.e., SSRs derived from ESTs or cDNA sequences). From the available 25,456 ESTs or cDNA sequences, we were able to develop 2,680 EST-SSRs (Huang 2008). These newly developed sorghum EST-SSR markers represent an additional resource for genetic mapping, comparative

genomics, as well as evaluation of colocation between QTLs and functionally associated markers in target species. Among these newly identified markers, a sub-set of 200 randomly selected EST-SSR markers was examined for the transferability to related cereal crops, showing their potential as molecular markers in maize, sugarcane, rice, wheat, and barley. EST-SSR markers offer the potential to cover the gene-rich regions of the entire nuclear genome and to fill gaps in linkage maps using comparative genomic information (Ramu et al. 2009).

With the availability of dense linkage maps of the sorghum genome, progress in the identification of genes or QTLs linked to plant resistance to diseases, insects, and abiotic stresses has been made. For example, greenbug aphid has been the most damaging insect pest of sorghum in the USA and often causes severe crop damage and significant economic loss. Using SSR markers, we have identified major QTLs conferring resistance to greenbug biotype E (Wu et al. 2007) as well as to greenbug biotype I (Wu and Huang 2008). A separate study reported that two genetic regions located on separate linkage groups were found to be associated with midge resistance and explained 12% and 15% of the total variation, respectively (Tao et al. 2003). At ICRISAT, mapping populations have been phenotyped and genotyped for sorghum shoot fly (296B × IS 18551 and BTx 623 × IS 18551), spotted stem borer, sorghum midge, and aphid (ICSV 745 × PB 15881-3). Genetic linkage maps based on these populations have been constructed to identify QTLs associated with resistance to these insects. Polymorphic simple sequence repeat (SSR) loci associated with resistance to shoot fly and the traits associated with resistance to this insect have been identified (Folkertsma et al. 2003; Dhillon et al. 2006e; Satish et al. 2009; Aruna et al. 2011). These QTLs are now being transferred into the locally adapted hybrid parental lines via SSR-based marker-assisted selection. QTLs associated with resistance to sorghum head bug (*Eurystylus oldi* Poppius) have also been identified (Deu et al. 2005). In addition, Satish et al. (2009) reported that QTL identified in this study correspond to QTL/genes for insect resistance at the syntenic maize genomic regions, suggesting the conservation of insect-resistance loci between these crops.

In general, it takes five to six generations to transfer a trait within a species into high-yielding, locally adapted cultivars through conventional breeding, and in this way, one has to evaluate a large number of progenies to be able to select the plants with the appropriate combination of traits. Fortunately, use of DNA markers for indirect selection offers great potential gains for quantitative traits with low heritability, as these are the most difficult characters to work with in the field using direct phenotypic selection. The effectiveness of a marker-assisted selection (MAS) can only be as good as the quality of the phenotypic data on which the development of the marker was based. It is hoped that MAS will allow rapid introgression of the resistance genes, and ultimately gene pyramiding, into the high-yielding varieties and hybrids. No doubt, all of the resultant information from the above-mentioned studies will facilitate the early selection of breeding lines through marker-assisted selection and cloning of the important resistance genes for sorghum improvement via the map-based cloning method.

7 Gene Expression Analysis and Gene Discovery

In addition to constitutive defenses, initiation and utilization of specific defense responses to attacking insect pests are important strategies for plant persistence and survival. Inducible defenses of plants consist of three steps: perception of herbivore, signal transduction, and biosynthesis of defensive products (Wu and Baldwin 2010). A wide range of inducible genes has been identified in plants based on endogenous chemical signals such as phytohormones, response to insect attack, or wounding. Chemically induced expression systems or "gene switches" enable the temporal, spatial, and quantitative control of genes introduced into crop plants, or those that are already present in the plants. The best-studied system utilizes pathogenesis-related protein-1a (*PR 1-a*) gene expression in tobacco (Uknes et al. 1993). The *PR 1-a* mRNA levels can also be induced by exogenous application of salicylic acid (Ward et al. 1991). Peptide hormones also induce production of proteinase inhibitors. Systemically induced responses are modified through synthesis and action of jasmonic acid via its lipid precursor, e.g., linoleic acid in tomato. Application of exogenous jasmonate induces the production of proteinase inhibitors. Enhanced resistance in transgenic rice plants by application of methyl jasmonate and abscisic acid has been observed (Xu et al. 1993).

The mechanisms of inducible plant defense responses are based on changes in gene expression. The rapid pace of advances into the molecular events of plant perception of pathogens has been particularly inspiring for the study of how plants perceive insect attack, although the same type of research is more advanced in the study of plant diseases. Microarray experiments for analyzing plant responses to insect attack have already shown promise for functional characterization of important processes such as plant defense. Zhu-Salzman et al. (2004) evaluated the transcriptional changes in a sorghum cultivar by comparing expression patterns of 672 cDNAs in the seedling tissues before and after infestation by greenbug or following treatment with defense signal components such as salicylic acid (SA) or methyl jasmonate (MJ). Their results indicated that activation of certain transcripts regulated exclusively by greenbug infestation was observed, and the expression patterns may represent unique signal transduction events independent of MJ- and SA-regulated pathways. More recently, we have examined the transcriptional changes in a parallel system, greenbug-resistant and -susceptible genotypes of sorghum, leading to detection of the abundance of the transcripts corresponding to 2,304 sorghum genes during the infestation by virulent greenbug biotype I (Park et al. 2006). The experiments showed comprehensive gene activation resulting from up-regulating, or activating existing defense pathways in sorghum seedlings in response to greenbug feeding. Among the induced genes identified in this study, 38 genes exhibited threefold or higher abundance in their expression, and 26 genes were significantly reduced. These cutting-edge technologies can enhance the understanding of plant defense mechanisms against insect pests, and also accelerate the identification of resistance genes or specific targets for improvement of plant resistance for integrated pest management in agriculture.

8 Transgenic Approaches of Insect Resistance in Sorghum

Plant biotechnology has become a promising tool for agricultural revolution, providing new solutions to age-old agricultural practices. Especially, significant advances in gene identification and gene transfer techniques allow the incorporation of beneficial genes for specific agronomic traits into diverse crop plants. Today these new tools enable plant breeders to design new varieties by installing desired foreign genes, such as insect- and disease-resistance genes, into existing commercial lines or elite breeding lines in a considerably short period of time. During the last decade, many of the world's most important crops (including wheat, maize, rice, soybean, and cotton) had already been engineered with increased resistance to insects and diseases (Sahrawat et al. 2003). Similar research has already been attempted in sorghum. Early research activities in sorghum transformation focused on developing gene transfer systems; thus successful genetic transformation systems were developed using either particle bombardment (Casas et al. 1993) or *Agrobacterium*-mediated transformation (Zhao et al. 2000). Though both *Agrobacterium*-mediated and particle bombardment transformation systems are successful in sorghum, the most effective method to date is *Agrobacterium*-based transformation, with a transformation efficiency of 2.1–4.5% (Zhao et al. 2000; Gao et al. 2005; Howe et al. 2006). Further information on the current status of sorghum transformation is provided elsewhere in this volume (Tejinder et al. 2012).

Today, sorghum producers face a major threat to their crops from insect pests worldwide. Given the wide host range of some of the insect pests, and low levels of resistance in the cultivated germplasm against major sorghum pests such as stem borers, head bugs, and armyworms, it will be highly desirable to combine conventional plant resistance with novel genes from other sources such as *Bacillus thuringiensis* (*Bt*) toxic proteins, protease inhibitors, or plant lectins. Streamlined sorghum transformation could help in production of transgenic plants with improved resistance to important insect pests. Sorghum plants having the *cry1Ac* gene have been developed under the control of a wound-inducible promoter from a maize protease inhibitor gene (*mpi*) for resistance to spotted stem borer, *C. partellus* (Seetharama et al. 2001; Harshavardhan et al. 2002; Girijashankar et al. 2005). Feeding by the neonate larvae of *C. partellus* on the leaf discs from transgenic plants was 60% lower compared to that on the nontransgenic control plants, and the weight gain by the larvae was reduced by 36%, which could be due to low levels of *Bt* protein expressed in sorghum leaf tissues, as it was 1–8 ng g^{-1} of fresh leaf tissue. The larval mortality was 40% more in the larvae fed on leaf discs from transgenic plants compared to that on the nontransgenic control plants. Thus, combining transgenic resistance to insects with the conventional plant resistance will make HPR an effective weapon for pest management in sorghum. More recently, Kosambo-Ayoo et al. (2011) reported that sorghum was transformed with the chitinases and chitosanases genes isolated from *Trichoderma harzianum*. Seedlings from a transgenic line were found to be significantly more tolerant to anthracnose than the parent wild

type. The above successful examples demonstrate that transgenic technology can pyramid resistance genes to multiply the effectiveness of beneficial genes in sorghum crops with multiple resistances to insect pests and diseases.

As the transformation systems are available in sorghum, now the speed of gene discovery is the bottleneck for molecular breeding in sorghum. It is noteworthy that some secondary plant metabolites such as flavonoids have been implicated in HPR to insects in sorghum. Many compounds of the flavonoid biosynthetic pathway accumulate in response to biotic and abiotic stresses (Heller and Forkman 1993). Genetic engineering offers the opportunity to change the metabolic pathways to increase the amounts of various flavonoids, which play an important role in HPR to insect pests (Zhuang et al. 2011). Biotechnology also offers the opportunity to increase the production of secondary metabolites in plants to increase the levels of resistance to insect pests or inhibit the production of toxic metabolites such as HCN in forage sorghum. Thus, combining transgenic resistance to insects with conventional plant resistance will make plant resistance an effective component for pest management in sorghum.

9 Summary

Sorghum has an advantage over other grain crops because it can withstand relatively harsh, hot, dry climates, but responds well to favorable production conditions and irrigation. Thus, sorghum has a reduced vulnerability to climate change, becoming more important in future agriculture as a drought-tolerant, fast-growing crop, which can thrive and yield relatively well even with high water scarcity within the shortened length of the growing seasons, as it matures before the depletion of soil moisture, thereby reducing the threat from dry spells (Dar 2009). In addition, sorghum, a leading and cost-effective bioenergy crop, is poised to have an important role in crop production worldwide as farmers are willing to grow sorghum as a cellulosic bioenergy crop in order to help meet the demands of renewable fuels to produce "next generation" fuels. However, insect pests will be a major limiting factor in sustainable production of sorghum in the future. Among the major insect pests that limit its production in various geographic areas are greenbug, midge, shoot fly, stem borer, and head worms.

HPR is one of the most economic means of controlling insect pests without the undesirable effects of pesticides. The utilization of natural resistance for pest management has been successful in the past and some of the resistances have been transferred into commercial cultivars and hybrids. Continuous research efforts on identification of new sources of resistance to the major pests of sorghum and breeding for these types of resistance are major research objectives, along with basic studies on insect–host plant relationships for a better understanding of the nature of resistance. In germplasm evaluation for pest resistance, sorghum researchers need to improve precision of screening and selection criteria for resistance to insect pests in many cases. Genetic resistance in sorghum, wherever available, should be combined

with other desirable plant characters, such as high yield, good quality, disease resistance, and should provide the basic foundation on which to build integrated pest management systems. A promising strategy for sorghum improvement should be based on gene pyramiding and development of cultivars with multiple resistances to insect pests and diseases. There is an urgent need to transfer various insect-resistance genes into CMS, maintainer, and restorer lines, so as to be able to develop hybrids with increased levels, and diverse mechanisms of resistance to target pests.

Despite progressing beyond basic research involving the development of transformation methods for introducing useful genes to the sorghum genome, gene transfer technology in sorghum is still at a juvenile stage. Protocols need to be optimized in order to develop simple procedures with improved transformation efficiency. Transgenic technology will certainly make it easier to transfer resistance genes from wild relatives or other sources and can assist in the production of agronomically desirable crops that have, for example, improved ability to defend themselves against insect pests.

Genomic technologies have aided tremendously in identifying loci or genomic regions associated with insect resistance in most crop plants including sorghum. These genomic tools and molecular markers not only promise to increase our knowledge of mechanisms underlying host resistance to insect pests in sorghum, but also facilitate the marker-assisted breeding in our future sorghum breeding programs. All above-mentioned research approaches promise to facilitate development of insect-resistant cultivars and hybrids of sorghum. However, the complexity in improving levels of pest resistance requires collaborative efforts between research institutions and the sorghum industry as well as international cooperation in utilization of emerging knowledge and technologies to enhance the global efforts in insect pest management.

References

Aruna C, Bhagwat VR, Madhusudhana R, Sharma V, Hussain T, Ghorade RB, Khandalkar HG, Audilakshmi S, Seetharama N (2011) Identification and validation of genomic regions that affect shoot fly resistance in sorghum [*Sorghum bicolor* (L.) Moench]. Theor Appl Genet 122:1617–1630

Bhattramakki D, Dong J, Chhabra AK, Hart G (2000) An integrated SSR and RFLP linkage map of *Sorghum bicolor* (L.) Moench. Genome 43:988–1002

Borad PK, Mittal VP (1983) Assessment of losses caused by pest complex to sorghum hybrid CSH 5. In: Krishnamurthy Rao BH, Murthy KSRK (eds) Crop losses due to insect pests. Entomological Society of India, Andhra Pradesh, India, pp 271–278

Bowers JE, Abbey C, Anderson S, Chang C, Draye X, Hoppe AH, Jessup R, Lemke C, Lennington J, Li Z, Lin Y-R, Liu S-C, Luo L, Marler BS, Ming R, Mitchell SE, Qiang D, Reischmann K, Schulze SR, Skinner DN, Wang Y-W, Kresovich S, Schertz KF, Paterson AH (2003) A high-density genetic recombination map of sequence-tagged sites for *Sorghum*, as a framework for comparative, structural and evolutionary genomics of tropical grains and grasses. Genetics 165:367–386

Burd JD, Porter DR (2006) Biotypic diversity in greenbug (Hemiptera, Aphididae), characterizing new virulence and host associations. J Econ Entomol 99:959–965

Casas AM, Kononowicz AK, Zehr UB, Tomes DT, Axtell JD, Butler LG, Bressan RA, Hasegawa PM (1993) Transgenic sorghum plants via microprojectile bombardment. Proc Natl Acad Sci (USA) 90:11212–11216

Chittenden LM, Schertz KF, Lin YR, Wing RA, Paterson AH (1994) A detailed RFLP map of *Sorghum bicolor* × *S. propinquum*, suitable for high-density mapping, suggests ancestral duplication of Sorghum chromosomes or chromosomal segments. Theor Appl Genet 87:925–933

Dar WD (2009) Winning the gamble against the monsoons. http://www.hindu.com/2009/07/05/stories/2009070555380900.htm

Deu M, Ratnadass MA, Hamada MA, Noyer JL, Diabate M, Chantereau J (2005) Quantitative trait loci for head-bug resistance in Sorghum. Afr J Biotechnol 4:247–250

Dhillon MK, Sharma HC, Naresh JS, Ram S, Pampapathy G (2006a) Influence of cytoplasmic male-sterility on different mechanisms of resistance in sorghum to shoot fly *Atherigona soccata*. J Econ Entomol 99(4):1452–1461

Dhillon MK, Sharma HC, Smith CM (2008) Implications of cytoplasmic male-sterility systems for development and deployment of pest resistant hybrids in cereals. CAB Rev Prospect Agric Vet Sci Nutrit Nat Res 3(068):1–16

Dhillon MK (2004) Effects of cytoplasmic male-sterility on expression of resistance to sorghum shoot fly, *Atherigona soccata* (Rondani) (Muscidae, Diptera). PhD thesis. Department of Entomology, Chaudhary Charan Singh Haryana Agricultural University, Hisar, Haryana, India, 382 pp

Dhillon MK, Sharma HC, Reddy BVS, Ram S, Naresh JS, Kai Z (2005) Relative susceptibility of different male-sterile cytoplasms in sorghum to shoot fly, *Atherigona soccata*. Euphytica 144:275–283

Dhillon MK, Sharma HC, Ram S, Naresh JS (2006b) Influence of cytoplasmic male-sterility on expression of physico-chemical traits associated with resistance to sorghum shoot fly, *Atherigona soccata*. SABRAO J Breed Genet 38:105–122

Dhillon MK, Sharma HC, Pampapathy G, Reddy BVS (2006c) Cytoplasmic male-sterility affects expression of resistance to shoot bug (*Peregrinus maidis*), sugarcane aphid (*Melanaphis sacchari*) and spotted stem borer (*Chilo partellus*). Intl Sorghum Millets Newslett 47:66–68

Dhillon MK, Sharma HC, Reddy BVS, Ram S, Naresh JS (2006d) Nature of gene action for resistance to sorghum shoot fly, *Atherigona soccata*. Crop Sci 46:1377–1383

Dhillon MK, Sharma HC, Folkertsma RT, Chandra S (2006e) Genetic divergence and molecular characterization of shoot fly-resistant and -susceptible parents and their hybrids. Euphytica 149:199–210

Eddleman BR, Chang CC, McCarl BA (1999) Economic benefits from grain sorghum variety improvement in the United States. In: Wiseman BR, Webster JA (eds) Economic, environmental, and social benefits of resistance in field Crops. Entomological Society of America, Lanham, MD, pp 17–44

Folkertsma RT, Sajjanar GM, Reddy BVS, Sharma HC, Hash CT (2003) Genetic mapping of QTL associated with sorghum shoot fly (*Atherigona soccata*) resistance in sorghum (*Sorghum bicolor*). In: Final abstracts guide, plant & animal genome XI, 11–15 Jan 2003. San Diego, CA, USA, p 42. http://www.intl-pag.org/11/abstracts/P5d_P462_XI.html

Gao Z, Xie X, Ling Y, Muthukrishnan S, Liang GH (2005) *Agrobacterium tumefaciens*-mediated sorghum transformation using a mannose selection system. Plant Biotechnol J 3:591–599

Girijashankar V, Sharma HC, Sharma KK, Sivarama PL, Royer M, Secundo BS, Lakshmi N, Seetharama N (2005) Development of transgenic sorghum for insect resistance against spotted stem borer, (*Chilo partellus*). Transgen Res (in press)

Harshavardhan D, Rani TS, Sharma HC, Arora R, Seetharama N (2002) Development and testing of *Bt* transgenic sorghum. In: International symposium on molecular approaches to improve crop productivity and quality, 22–24 May 2002, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India

Harvey TL, Hackerott HL (1969) Recognition of a greenbug biotype injurious to sorghum. J Econ Entomol 62:776–779

Heller W, Forkman G (1993) Biosynthesis of flavonoids. In: Harborne JB (ed) The flavonoids, advances in research since 1986. Chapman and Hall, London

Henzell RG, Brengman RL, Page FD (1980) Transference of sorghum midge resistance in to agronomically acceptable lines. In: Proc. 1st Australian Agronomy Conference. April 1980, Lawes, Queensland

Henzell RG, Jordan DR (2009) Grain sorghum breeding. In: Carena MJ (ed) Cereals. Springer Science, New York, pp 183–197

Howe A, Sato S, Dweikat I, Fromm M, Clemente T (2006) Rapid and reproducible *Agrobacterium*-mediated transformation of sorghum. Plant Cell Rep 25:784–791

Huang Y (2004) Examining plant defense responses to greenbug attack in sorghum using DNA microarray technology. Intl Sorghum Millets Newslett 44:72–74

Huang Y (2006) Evaluating sorghum germplasm for resistance to greenbug (*Schizaphis graminum*) biotype I. Intl Sorghum Millets Newslett 47:72–74

Huang Y (2007) Phloem feeding regulates the plant defense pathways responding to both aphid infestation and pathogen infection. In: Zhi-hong Xu et al (eds) Biotechnology and sustainable agriculture 2006 and beyond. Springer, New York, pp 215–219

Huang Y (2008) Development of EST-SSR markers for sorghum and their transferability among cereal species. In: Proc. Intl. Plant & Animal Genome Conference. 12–26 Jan 2008, San Diego, CA, pp 148

Huang Y (2011) Improvement of crop protection against insect pest using worldwide germplasm collection and genomics-based approaches. Plant Genet Resour Charact Utiliz 9:317–320

Hulbert SH, Richter TE, Axtell JD, Bennetzen JL (1990) Genetic mapping and characterization of sorghum and related crops by means of maize DNA probes. Proc Natl Acad Sci (USA) 87: 4251–4255

ICRISAT (1992) Annual report 1991. International Crop research Institute for Semi-arid Tropics. Patancheru, Andhra Pradesh, India

ICRISAT (International Crops Research Institute for the Semi-Arid Tropics) (1989) International workshop on sorghum stem borers, 17–20 Nov 1987, ICRISAT Center, Patancheru, Andhra Pradesh, India

Johnson JW, Rosenow DT, Teetes GL (1973) Resistance to the sorghum midge in converted exotic sorghum cultivars. Crop Sci 13:754–755

Johnson JW (1977) Status of breeding for midge resistance. 10th biennial grain sorghum research and utilization conference, 2–4 Mar 1977, Grain Sorghum Producers Association, Wichita, KS

Kim J-S, Klein PE, Klein RR, Price HJ, Mullet JE, Stelly DM (2005) Chromosome identification and nomenclature of *Sorghum bicolor*. Genetics 169:1169–1173

Kimber CT, Dahlberg JA, Kresovich S (2012) The gene pool of *Sorghum bicolor* and its improvement. In: Paterson AH (ed) Genomics of the saccharinae. Springer, New York, pp 23–41

Klein PE, Klein RR, Cartinhour SW, Ulanch PE, Dong J, Obert JA, Morishige DT, Schlueter SD, Childs KL, Ale M, Mullet JE (2000) A high-throughput AFLP-based method for constructing integrated genetic and physical maps. Progress toward a sorghum genome map. Genome Res 10:789–807

Kosambo-Ayoo LM, Bader M, Loerz H, Becker D (2011) Transgenic sorghum (*Sorghum bicolor L. Moench*) developed by transformation with *chitinase* and *chitosanase* genes from *Trichoderma harzianum* expresses tolerance to anthracnose. Afr J Biotechnol 10:3659–3670

Kumar H (1993) Responses of *Chilo partellus* (Lepidoptera, Pyralidae) and *Busseola fusca* (Lepidoptera, Noctuidae) to hybrids of a resistant and a susceptible maize. J Econ Entomol 86:962–968

Kumar H, Mihm JA (1996) Resistance in maize hybrids and inbreds to first-generation southwestern corn borer, *Diatraea grandiosella* (Dyar) and sugarcane borer, *Diatraea saccharalis* Fabricius. Crop Prot 15:311–317

Kumari AP, Sharma HC, Reddy DDR (2000) Components of resistance to sorghum head bug, *Calocoris angustatus*. Crop Prot 19:385–392

Moran JL, Rooney WL (2003) Effect of cytoplasm on the agronomic performance of grain sorghum hybrids. Crop Sci 43:777–781

Mote UN (1984) Sorghum species resistant to shoot fly. Indian J Entomol 46:241–243

Painter H (1951) Insect resistance in host plants. Macmillan, New York, p 520

Park SJ, Huang Y, Ayoubi P (2006) Identification of expression profiles of sorghum genes in response to greenbug phloem-feeding using cDNA subtraction and microarray analysis. Planta 223:932–947

Pereira MG, Lee M, Bramel-Cox P, Woodman W, Doebley J, Whitkus J (1994) Construction of an RFLP map in sorghum and comparative mapping in maize. Genome 37:236–243

Peterson GC, Reddy BVS, Youm O, Teetes GL, Lambright L (1997). Breeding for resistance to foliar- and stem-feeding insects of sorghum and pearl millet. In: Proceedings of the International Conference on Genetic Improvement of Sorghum and Pearl Millet. INTSORMIL, Publ. 97-5, pp 281–302

Ramu P, Kassahun B, Senthilvel S, Ashok KC, Jayashree B, Folkertsma RT, Ananda Reddy L, Kuruvinashetti MS, Haussmann BIG, Hash CT (2009) Exploiting rice–sorghum synteny for targeted development of EST-SSRs to enrich the sorghum genetic linkage map. Theor Appl Genet 119:1193–1204

Reddy BVS, Stenhouse JW (1994) Improving post-rainy season sorghum, a case study for landrace hybrid approach. An invited paper presented at All India co-ordinated sorghum improvement project (A ICSIP) workshop held at Pantnagar, UP, 18–20 April

Rooney WL (2004) Sorghum improvement, integrating traditional and new technology to produce improved genotypes. Adv Agron 83:37–109

Ross WM, Kofoid KD (1979) Effect of non-milo cytoplasms on the agronomic performance of sorghum. Crop Sci 19:267–270

Sahrawat AK, Becker D, Lütticke S, Lörz H (2003) Genetic improvement of wheat via alien gene transfer, an assessment. Plant Sci 165:1147–1168

Satish K, Srinivas G, Madhusudhana R, Padmaja PG, Nagaraja Reddy R, Murali Mohan S, Seetharama N (2009) Identification of quantitative trait loci for resistance to shoot fly in sorghum [Sorghum bicolor (L.) Moench]. Theor Appl Genet 119:1425–1439

Seetharama N, Mythili PK, Rani TS, Harshavardhan D, Ranjani A, Sharma HC (2001) Tissue culture and alien gene transfer in sorghum. In: Singh RP, Jaiwal PK (eds) Improvement of food crops. Sci-Tech Publishing Company, Houstan, TX, pp 235–266

Schertz KF (1994) Male-sterility in sorghum: its characteristics and importance. In: Witcombe JR, Duncan RR (eds) Use of molecular markers in sorghum and pearl millet breeding for developing countries. In: Proceedings of the international conference on genetic improvement of an overseas development administration (ODA) plant sciences research conference, 29 March–1 April 1993, Norwich, UK, ODA, UK, pp 35–37

Sharma HC (1993) Host plant resistance to insects in sorghum and its role in integrated pest management. Crop Prot 12:11–34

Sharma HC (2001) Cytoplasmic male-sterility and source of pollen influence the expression of resistance to sorghum midge, *Stenodiplosis sorghicola*. Euphytica 122:391–395

Sharma HC, Abraham CV, Vidyasagar P, Stenhouse JW (1996) Gene action for resistance in sorghum to midge, *Contarinia sorghicola*. Crop Sci 36:259–265

Sharma HC, Dhillon MK, Naresh JS, Ram S, Pampapathy G, Reddy BVS (2004) Influence of cytoplasmic male-sterility on the expression of resistance to insects in sorghum. In: Fisher T, Turner N, Angus J, McIntyre L, Robertson M, Borrell A, Llyod D. (eds) Fourth international crop science congress, 25 September–October 1, 2004. Brisbane, Queensland, Australia 2007

Sharma HC, Dhillon MK, Reddy BVS (2006) Expression of resistance to sorghum shoot fly in F_1 hybrids involving shoot fly resistant and susceptible cytoplasmic male-sterile and restorer lines of sorghum. Plant Breed 125:473–477

Sharma HC, Franzmann BA (2001) Host plant preference and oviposition responses of the sorghum midge, *Stenodiplosis sorghicola* (Coquillett) (Dipt., *Cecidomyiidae*) towards wild relatives of sorghum. J Appl Entomol 125:109–114

Sharma HC, Nwanze KF (1997) Mechanisms of resistance to insects and their usefulness in sorghum improvement. Information bulletin no. 55. International Crop Research Institute for the Semi-Arid Tropics, Patancheru, Andhra Pradesh, India, 51 pp

Sharma HC, Reddy BVS, Dhillon MK, Venkateswaran K, Singh BU, Pampapathy G, Folkertsma RT, Hash CT, Sharma KK (2005) Host plant resistance to insects in sorghum, present status and need for future research. Intl Sorghum Millets Newslett 46:36–43

Sharma HC, Taneja SL, Kameswara Rao N, Prasada Rao KE (2003) Evaluation of sorghum germplasm for resistance to insect pests. Information bulletin no. 63. Patancheru, Andhra Pradesh, India, International Crops Research Institute for the Semi-Arid Tropics (CRISAT). 184 pp

Sharma HC, Taneja SL, Leuschner K, Nwanze KF (1992) Techniques to screen sorghums for resistance to insects. Information bulletin no. 32. International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Andhra Pradesh, India, 48 pp

Sharma HC, Vidyasagar P, Leuschner K (1988a) Field screening for resistance to sorghum midge (Diptera, Cecidomyiidae). J Econ Entomol 81:327–334

Sharma HC, Vidyasagar P, Leuschner K (1988b) No-choice cage technique to screen for resistance to sorghum midge (Diptera, Cecidomyiidae). J Econ Entomol 81:415–422

Sharma HC, Vidyasagar P, Abraham CV, Nwanze KF (1994) Effect of cytoplasmic male-sterility in sorghum on host plant interaction with sorghum midge, *Contarinia sorghicola*. Euphytica 74:35–39

Tao YZ, Hardy A, Drenth J, Henzell RG, Franzmann BA, Jordan DR, Butler DG, McIntyre CL (2003) Identifications of two different mechanisms for sorghum midge resistance through QTL mapping. Theor Appl Genet 107:116–122

Tatum LA (1971) The southern corn leaf blight epidemic. Science 171:1113–1116

Teetes GL, Pendleton BB (2000) Insect pests of sorghum. In: Smith CW, Frederiksen RA (eds) Sorghum, origin, history, technology, and production. Wiley, New York, pp 443–495

Tejinder K, Howe A, Sato S, Dweikat I, Clemente T (2012) Sorghum transformation: overview and utility. In: Paterson AH (ed) Genomics of the saccharinae. Springer, New York, pp 205–221

Tryon H (1895) The insect enemies of cereals belonging to the genus Cecidomyia. Trans Nat Hist Soc Queensland 1:80–83

Uknes S, Dincher S, Friedrich L, Negrotto D, Williams S, Thompson-Taylor H, Potter S, Ward E, Ryals J (1993) Regulation of pathogenesis-related Protein-1a gene expression in tobacco. Plant Cell 5:159–169

van den Berg J, van Rensburg GDJ, van der Westhiizen MC (1994) Host-plant resistance and chemical control of Chilo partellus (Swinhoe) and Busseola fusca (Fuller) in an integrated pest management system on grain sorghum. Crop Prot 13:308–310

Venkateswaran K (2003) Diversity analysis and identification of sources of resistance to downy mildew, shoot fly and stem borer in wild sorghums. Ph.D. thesis. Hyderabad, Andhra Pradesh, India, Department of Genetics, Osmania University

Venkateswaran K, Sharma HC, Manohar Rao D, Varaprasad KS, Bramel PJ (2009) Wild relatives of sorghum as sources of resistance to sorghum shoot fly, *Atherigona soccata*. Plant Breed 128:137–142

Ward ER, Uknes SJ, Williams SC, Dincher SS, Wiederhold DL, Aleander DC, Ahl-Goy P, Metraux JP, Ryals JA (1991) Coordinate gene activity in response to agents that induce systemic acquired resistance. Plant Cell 3:1085–1094

Wu J, Baldwin IT (2010) New insights into plant responses to the attack from insect herbivores. Ann Rev Genet 44:1–24

Wu Y, Huang Y (2007) An SSR genetic map of *Sorghum bicolor* (L.) Moench and its comparison to a published genetic map. Genome 50:84–89

Wu Y, Huang Y (2008) Molecular mapping of QTLs for resistance to the greenbug *Schizaphis graminum* (Rondani) in *Sorghum bicolor* (Moench). Theor Appl Genet 117:117–124

Wu Y, Huang Y, Porter DR, Tauer CG, Hollaway L (2007) Identification of a major QTL conditioning resistance to greenbug biotype E in Sorghum PI 550610 using SSR markers. J Econ Entomol 100:1672–1678

Xu D, McElroy D, Thoraburg RW, Wu R (1993) Systemic induction of a potato pin 2 promoter by wounding methyl jasmonate and abscisic acid in transgenic rice plants. Plant Mol Biol 22:573–588

Yang W, de Oliveira AC, Godwin I, Schertz K, Bennetzen JL (1996) Comparison of DNA marker technologies in characterizing plant genome diversity, variability in Chinese sorghums. Crop Sci 36:1669–1676

Young WR, Teetes GL (1977) Sorghum entomology. Ann Rev Entomol 22:193–218

Zhuang X, Köllner TG, Zhao N, Li G, Jiang Y, Zhu L, Ma J, Degenhardt J, Chen F (2011) Dynamic evolution of herbivore-induced sesquiterpene biosynthesis in sorghum and related grass crops. Plant J 69:70–80

Zhao ZY, Cai T, Tagliani L, Miller M, Wang N, Pang H, Rudert M, Schroeder S, Hondred D, Seltzer J, Pierce D (2000) Agrobacterium-mediated sorghum transformation. Plant Mol Biol 44:789–798

Zhu-Salzman K, Salzman RA, Ahn J-E, Koiwa H (2004) Transcriptional regulation of sorghum defense determinants against a phloem-feeding aphid. Plant Physiol 134:420–431

Chapter 17
Genetic Enhancement of Sorghum for Biomass Utilization

Wilfred Vermerris and Ana Saballos

Abstract Biomass produced from sorghum can be utilized as forage and silage to feed ruminant animals and as feedstock for biofuels and bio-based products. The efficiency of biomass utilization is a function of biomass composition and plant architecture. This chapter provides a description of the cell wall polymers that make up the bulk of sorghum biomass, along with information on the genes involved in their biosynthesis. The close evolutionary relationships among the grasses makes it possible to infer gene function across species. Newly developed genomics and bioinformatics resources offer exciting opportunities for the genetic enhancement of sorghum as a biomass crop.

Keywords Biomass • Bioenergy • Biofuels • Cell wall • Forage • Near infrared spectroscopy • Pyrolysis • Silage

1 Introduction: Utilization of Sorghum Biomass

The focus of this chapter is on the use of the vegetative parts of the sorghum plant (*Sorghum bicolor* (L.) Moench) and on ways to improve the utilization of these parts of the plant through genetics, genomics, and plant breeding. The vegetative aerial

W. Vermerris (✉)
Genetics Institute and Agronomy Department, University of Florida, 2033 Mowry Road, Gainesville, FL 32610-3610, USA

Department of Agricultural & Biological Engineering and Laboratory of Renewable Resources Engineering, Purdue University, West Lafayette, IN, USA
e-mail: wev@ufl.edu

A. Saballos
Genetics Institute and Agronomy Department, University of Florida, 2033 Mowry Road, Gainesville, FL 32610-3610, USA
e-mail: saballoa@ufl.edu

A.H. Paterson (ed.), *Genomics of the Saccharinae*, Plant Genetics and Genomics: Crops and Models 11, DOI 10.1007/978-1-4419-5947-8_17,

parts of the plant are often referred to as biomass, which is essentially comprised of the stem, tillers, and leaves. Biomass can be harvested for different purposes, namely as nutrition for ruminant animals in the form of forage and silage or as a feedstock for bioenergy, biofuels, or bio-based products.

Sorghum biomass can be fed to (ruminant) animals, either as green chop (harvested and cut pre-booting), hay (dried plants) or by letting the animals graze the plants in the field. Sudangrass and sorghum × sudangrass hybrids are attractive for grazing because sudangrass has thinner stems, narrower leaves, and more tillers than "regular" sorghum; the hybrid is intermediate in its appearance. Sudangrass is considered a separate subspecies, *Sorghum bicolor* (L.) Moench ssp. *drummondii* ((Nees ex Steud.) de Wet & Harlan) (cf. USDA Natural Resources Conservation Service; http://plants.usda.gov), but has also been considered a member of the Working Group Sudanense of *Sorghum bicolor* ssp. *bicolor* (Dahlberg 2000). Silage is produced by chopping the entire above-ground part of the plant (including the grain), typically harvested at soft-dough stage. The chopped parts are preserved by microbial conversion of soluble sugars into low-molecular weight organic acids that lower the pH and prevent rot and decay. Either naturally occurring microorganisms or commercially available inoculants can be used. Silage is typically stored over the winter in so-called bunkers. Portions of the silage are removed and fed to the animals as needed. Silage tends to have a higher nutritional quality than forage because of the presence of starch (Smith and Frederiksen 2000). Sorghum silage competes with corn silage, which tends to have a higher net energy for lactation per hectare as long as water supplies during the growing season are not limited (Miron et al. 2007; Marsalis et al. 2010).

Biomass for bioenergy, biofuels, and bio-based products can be collected as residues after the grain harvest or can be produced from a dedicated biomass crop, whereby the biomass itself is the main product (Rooney et al. 2007; Carpita and McCann 2008). Since sorghum requires less water and fertilizer and grows well on poor soils and in dry climates, dedicated bioenergy sorghums intended for large-scale production are under active development. Due to the lack of a single conversion process (see Sect. 5), however, clear targets as to what constitutes the optimal biomass composition have yet to be defined.

2 Composition of Sorghum Biomass

2.1 The Sorghum Cell Wall

Sorghum biomass is primarily composed of plant cell walls. The cell wall is a complex matrix formed by several different polymers that together form a functional structure. The primary cell wall is formed during the late stages of cell division, when the phragmoplast, consisting of microtubules, microfilaments, and endoplasmic reticulum, guides the deposition of vesicles containing cell wall material to form a cell plate

rich in pectin. This layer forms the middle lamella onto which other cell wall components are deposited by both of the newly formed cells. A secondary cell wall is deposited in specialized cell types, specifically the water-conducting xylem cells and the structural support-providing sclerenchyma cells. The secondary cell wall is deposited in between the cell membrane and the primary wall. Cells that contain a secondary cell wall typically undergo programmed cell death, during which the cell content is purged (lysis) and a hollow tube is formed. The secondary wall is then exposed to the lumen side of this newly formed tube.

Plant cell walls can be classified based on the architecture of the primary wall (Carpita and Gibeaut 1993). The Type I wall is common in dicots and non-commelinoid monocots and contains a network of cellulose microfibrils held in place by the hemicellulosic polysaccharide xyloglucan, and embedded in a matrix formed by pectins. Cell wall proteins provide additional structural support. The Type II cell wall is found in commelinoid monocots, which include the grasses. Cellulose (**1**; Fig. 17.1) is crosslinked by the hemicellulosic polysaccharide glucuronoarabinoxylan

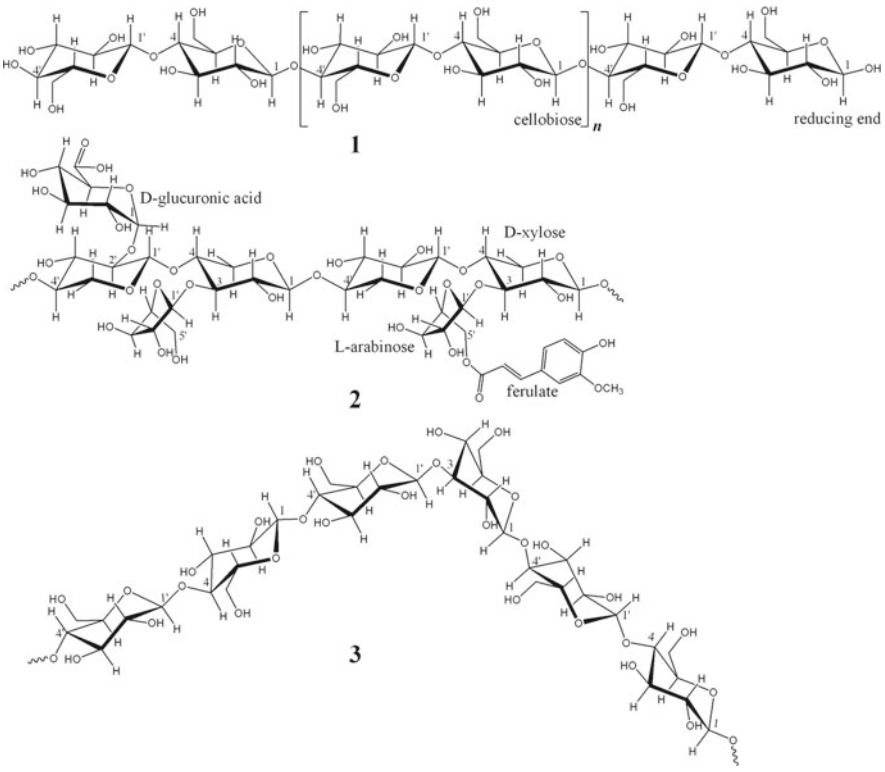

Fig. 17.1 Structural fragments of the most abundant cell wall polysaccharides in sorghum. Cellobiose consists of two β-1,4-linked D-glucopyranose residues and is the repeat unit of cellulose (**1**). GAX (**2**), the main hemicellulosic polysaccharide, consists of a xylan backbone with arabinose, glucuronic acid, and ferulic acid substitutions. Mixed-linkage β-glucans (**3**) consist of β-(1,4)-linked cellotriose and cellotetraose units connected via β-(1,3)-linkages

(GAX; **2**). The Type II wall has a lower pectin and protein content, and instead contains hydroxycinnamic acids and a mixed-linkage β-D-glucan polymer (**3**), which is hydrolyzed when the primary wall stops growing (Carpita 1996).

Cellulose (**1**) consists of linear strands of β-1,4-linked D-glucopyranose molecules that are rotated 180° relative to each other. A cellulose microfibril typically consists of 36 of these glucan strands held together via hydrogen bonds. Each strand consists of several thousand glucose residues. The individual glucan strands do not begin and end at the same place, enabling the formation of microfibrils that can be several hundred micrometers long. Cellulose is synthesized by rosettes of cellulose synthases (CesAs) located in the cell membrane. Several excellent reviews on cellulose and CesAs have been written by Zugenmaier (2001), Saxena and Brown (2005), Ding and Himmel (2006), and Somerville (2006), although their focus is not restricted to sorghum.

GAX (**2**) consists of a linear chain of xylose residues substituted with α-L-arabinose and β-D-glucuronic acid residues. The arabinose residues are esterified to xylose at the *O*-3 position, whereas the glucuronic acid residues are esterified at the *O*-2 position. Ferulic acid is esterified to a subset (approximately 1 in 50) of the arabinose residues. Using cell cultures of tall fescue grass, Myton and Fry (1994) showed that the feruloylation of arabinose occurs prior to incorporation into GAX. The assumption is that this mechanism occurs similarly in all grasses, including sorghum.

As a member of the Poales, sorghum also transiently forms mixed-linkage β-D-glucans (**3**). This hemicellulosic polysaccharide contains β-(1,4)-linked cellotriose and cellotetraose units, consisting of three and four D-glucose residues, respectively, that are connected via β-(1,3)-linkages. As a result of this configuration, the polymer has a corkscrew-like shape. The ratio of cellotriose to cellotetraose units varies by species (Fincher 2009).

The secondary cell wall contains up to 25 % of lignin, a phenolic polymer formed from the oxidative coupling of the monolignols *p*-coumaryl alcohol (**4**; Fig. 17.2), coniferyl alcohol (**5**), and sinapyl alcohol (**6**). Once incorporated into the lignin polymer, structures derived from these monolignols are referred to as *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) residues. Sinapyl alcohol does not readily participate in the crosslinking reaction. Its incorporation into lignin is mediated by the formation of *p*-coumaroylsinapate (**7**; Hatfield et al. 2008), which explains the high content of *p*-coumarate in the cell walls of grasses. The presence of ferulic acid enables the crosslinking of different GAX molecules through diferulate bridges involving different linkages (e.g., **9**; Marita et al. 2003). In addition, because of its ability to form peroxidase-mediated radicals just like the monolignols, ferulic acid esterified to GAX is thought to serve as a nucleation site for the deposition of lignin (Ralph et al. 1995, 2004a). While most of the lignin is present in the secondary cell walls, it is noteworthy that as grasses mature, lignin is also deposited in many of the primary cells walls of the stem parenchyma. Detailed reviews of monolignol biosynthesis and lignin polymerization have been written by Humphreys and Chapple (2002), Ralph et al. (2004c), Vermerris and Nicholson (2006), and Vanholme et al. (2008).

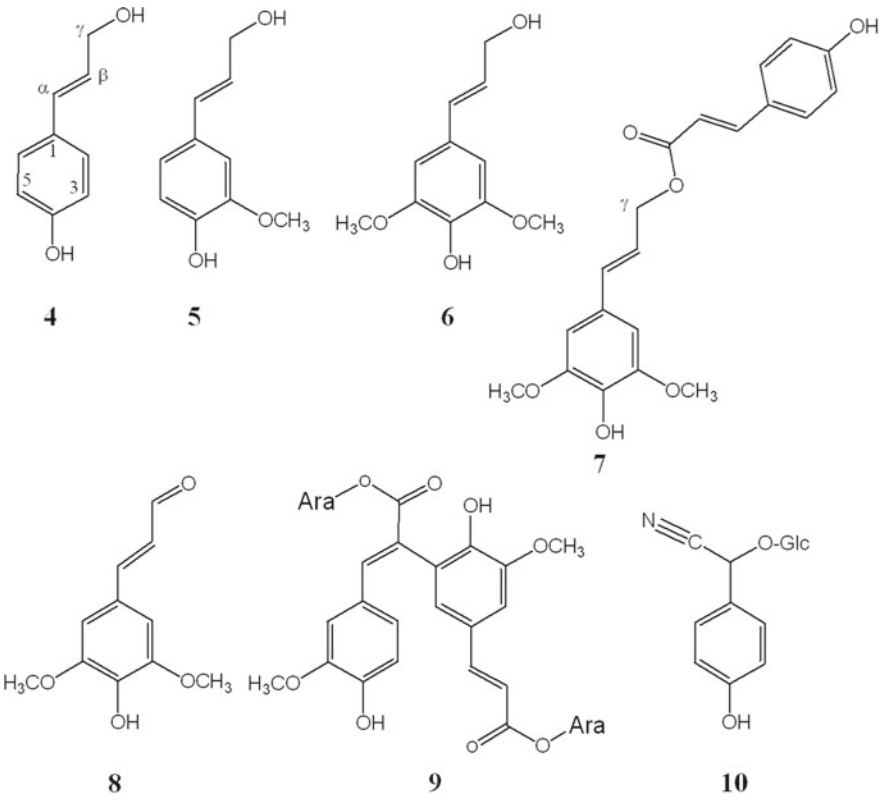

Fig. 17.2 The monolignols *p*-coumaryl alcohol (**4**), coniferyl alcohol (**5**), and sinapyl alcohol (**6**) can undergo oxidative coupling to form the cell wall polymer lignin. The incorporation of sinapyl alcohol is enhanced by the formation of its *p*-coumaroyl ester (**7**). Sinapyl aldehyde (**8**), the precursor of **6**, is the preferred substrate of the main lignification-related cinnamyl alcohol dehydrogenase in sorghum, encoded by the *Brown midrib6* gene. Ferulic acid residues esterified to arabinose residues in GAX can form diferulate bridges that enable crosslinking of different GAX polymers. Structure **9** represents the most common diferulate structure involving an 8–5′ linkage. Dhurrin (**10**) is a cyanogenic glucoside with antinutritional qualities that can accumulate in sorghum stems and leaves

Even though biomass is composed primarily of plant cell walls, the exact chemical composition of biomass varies with the genotype, the age at which the plants are harvested, and the environment. These factors define the relative proportion of stem and leaf tissue and the proportion of vascular tissue relative to nonvascular tissue, the relative proportions of the cell wall constituents, and their level of degradation as the plants senesce and are exposed to cell-wall degrading microorganisms. Despite this variation, generic sorghum biomass, as it is currently available, is composed of approximately 40 % cellulose, 20–25 % hemicellulosic polysaccharides (mostly GAX), 20 % lignin, 2.5 % hydroxycinnamic acids, 3–5 % pectin, 3–5 % protein,

5–8 % minerals, and 5 % starch (Rooney et al. 2007; Corredor et al. 2008; Li et al. 2010; Liu et al. 2010).

Detailed background information on various aspects of cell wall biogenesis and cell wall composition can be found in Carpita (1996), Ralph et al. (2004c), Vermerris and Nicholson (2006), and Knox (2008).

2.2 Assaying Sorghum Forage Quality

Researchers with a focus on forage quality have a choice of methods to determine the digestibility of the samples of interest. One option is to perform feeding studies with lactating cows (Miron et al. 2007; Marsalis et al. 2010), but the expense and time involved in such studies precludes them from being used on a routine basis. Tilley and Terry (1963) devised an in vitro digestibility method based on rumen fluid obtained from fistulated cows that enabled the processing of samples in the laboratory. In order to bypass the need for rumen fluid, Van Soest (1967) designed a series of analyses to determine the content of neutral detergent fiber (NDF), acid detergent fiber (ADF) and acid detergent lignin (ADL) in the forage samples. NDF refers to the total cell wall content (cellulose, hemicellulosic polysaccharides, pectins, lignin) and is determined gravimetrically after extracting starch, lipids, soluble sugars, and soluble phenolics from ground plant biomass in a solution containing sodium dodecyl sulfate (SDS). ADF represents the cellulose and lignin fraction of the NDF and is determined after subjecting the dried NDF to hydrolysis in 0.5 M sulfuric acid, which removes the hemicellulosic polysaccharides. ADL is determined after subjecting the cellulose in the ADF to hydrolysis in 12 M sulfuric acid. The insoluble residue—consisting of lignin and minerals—is collected on a filter, rinsed, dried, and weighed. The filter is then heated in a muffle furnace (450 °C, 5 h), which removes the lignin and leaves the minerals behind as an ash. ADL is calculated as the difference in the weight before and after heating. The different cell wall constituents can be calculated through subtraction: The content of hemicellulosic polysaccharides is calculated by subtracting ADF from NDF, whereas the cellulose content is calculated by subtracting ADL from ADF.

Additional methods to predict digestibility without having to obtain rumen fluid have been developed and include enzyme-based methods (De Boever et al. 1988) and predictions with the French acronyms DINAG and DINAGZ (Argillier et al. 1995; Fontaine et al. 2003). A detailed review of the different assays and the way they have been used in forage breeding programs is provided by Barrière et al. (2003).

A detailed method for cell wall analysis of forage samples is referred to as the "Uppsala method," named after the Swedish city where this method was developed (Theander et al. 1995). This procedure starts out with starch-free alcohol-insoluble residues prepared from the harvested biomass sample. Cell wall polysaccharides are then broken down to their corresponding monosaccharides following a two-stage hydrolysis in sulfuric acid. The neutral sugars (from cellulose and hemicellulosic polysaccharides) are derivatized to alditol acetates and quantified by gas chromatography, whereas the acid sugars from pectins are quantified with a

colorimetric method. Klason lignin is defined as the ash-free insoluble residue remaining after the acid hydrolysis. The Uppsala method provides detailed information on the composition of the cell wall, but it requires considerable investments in both equipment and time.

3 Cell Wall Biosynthetic Genes of Sorghum

3.1 Cellulose Synthases

Cellulose is synthesized by CesA, a membrane-bound enzyme that uses UDP-D-glucose as a substrate to generate long glucan strands. In plant cell walls cellulose is typically synthesized by a so-called terminal complex consisting of six clusters of six CesA subunits (Mueller and Brown 1980). Each CesA is responsible for synthesizing one glucan strand, so that a terminal complex will produce 36 glucan strands, which together form the cellulose microfibril (Saxena and Brown 2005; Somerville 2006). Seven microfibrils can form a macrofibril (Ding and Himmel 2006).

Each group of six subunits contains three pairs of different CesAs. The identity of the subunits differs between the primary and secondary cell wall. After the cloning of the first plant *cellulose synthase* gene from cotton (Pear et al. 1996), many of the studies that followed were performed in Arabidopsis. Arabidopsis contains ten *CesA* genes. The three subunits in the primary wall were shown to be CesA1, CesA3, and CesA6 (Arioli et al. 1998; Fagard et al. 2000; Scheible et al. 2001; Desprez et al. 2002), whereas in the secondary wall they are CesA4, CesA7, and CesA8 (Taylor et al. 1999, 2000, 2003). There appears to be some functional redundancy among the subunits (Desprez et al. 2007) and there is also evidence for developmental variation in the composition of the rosette structure. A similar but more complex situation occurs in maize (Holland et al. 2000; Appenzeller et al. 2004). Prior to the release of the maize genome sequence (Schnable et al. 2009), the most detailed study included 12 *ZmCesA* genes, which included three *ZmCesA* genes involved in secondary cell wall synthesis (*ZmCesA10-12*) and three clusters each containing three *ZmCesA* genes involved in primary cell wall synthesis. These nine genes appeared to be expressed differentially in different tissues and in different developmental stages (Appenzeller et al. 2004).

With the release of the maize genome sequence a total of 20 *ZmCesA* genes were reported (Schnable et al. 2009). These 20 *CesA* genes are reflective of the whole-genome duplication maize, underwent 5–12 MYA, and distinguish maize from rice and sorghum, as well as Arabidopsis with ten "ancestral" *CesA* genes each. The fact that maize appears to have retained all the ten duplicate copies of the *CesA* genes was reported as a somewhat unusual case, because on a genome-wide basis, only 25 % of the duplicated maize genes were retained, with a significant bias towards genes encoding transcription factors (Schnable et al. 2009). The spatio-temporal expression of the *CesA* genes not studied by Appenzeller et al. (2004) is still under investigation (Dr. Bryan Penning, personal communication), but based on currently

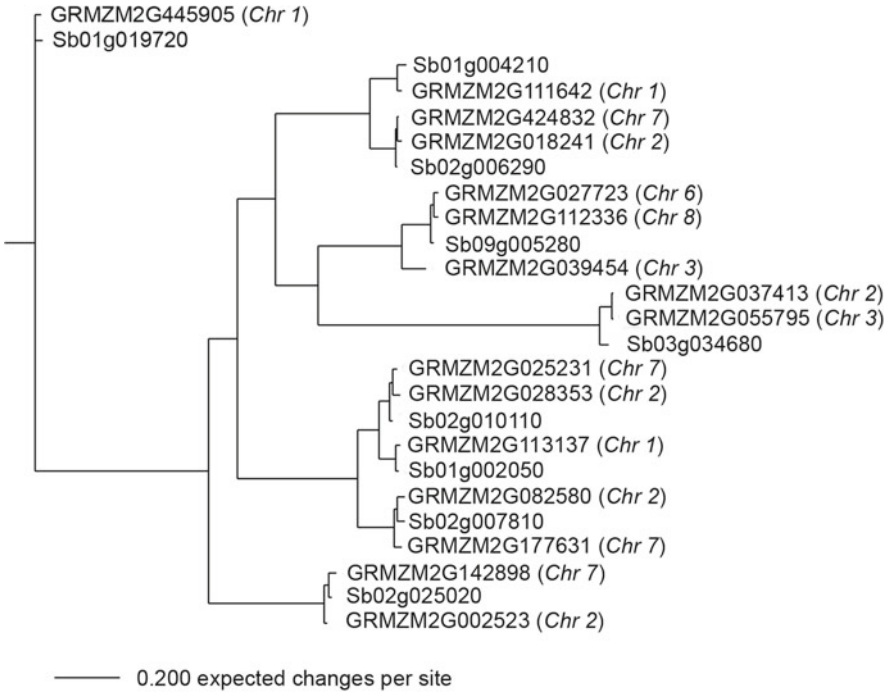

Fig. 17.3 The evolutionary relationship between sorghum and maize CesA proteins. Amino acid sequences were deduced from genome sequence and expression data for both species (Paterson et al. 2009; Schnable et al. 2009). The maize accession numbers were kindly provided by Drs. Nick Carpita and Bryan Penning (Purdue University) and correspond to data published by Schnable et al. (2009). The maize chromosomes on which individual *ZmCesA* genes reside are indicated in *parentheses*. The sorghum chromosome is indicated in the accession by the number following "Sb." The sequences were aligned using CLUSTALW (Thompson et al. 1994), followed by a Bayesian phylogenetic analysis using Mr.Bayes (http://mrbayes.csit.fsu.edu/index.php; Huelsenbeck and Ronquist 2001), with the amino acid model set to pr = mixed, parameters lset rates = gamma, number of generations = 27,000 with sampling every 100 generations (Ronquist et al. 2005). The output file was visualized using Phylodendron (http://iubio.bio.indiana.edu/treeapp/treeprint-form.html)

available expression data (EST reported in the genome databases), not all *CesA* genes appear to be expressed (e.g., Sb10g023430 and GRMZM2G104092). Furthermore, some of the reported *ZmCesA* genes may be duplicates to be resolved through further annotation. Figure 17.3 displays the phylogenetic relationships among the proteins encoded by nine expressed sorghum *SbCesA* genes (based on ESTs) and 15 unique, expressed maize *ZmCesA* orthologs. In six out of nine cases a sorghum *CesA* gene is matched by two maize orthologs. This tree is consistent with known evolutionary relationships among sorghum and maize chromosomes: sorghum chromosome 1 and maize chromosomes 1 and 9; sorghum chromosome 2 and maize chromosomes 2 and 7; sorghum chromosome 3 and maize chromosomes

3 (and 8); and sorghum chromosome 9 and maize chromosomes 3, 6, and 8 (Wilson et al. 1999; Kim et al. 2005).

3.2 Genes Involved in the Synthesis of Hemicellulosic Polysaccharides

The backbones of some of the hemicellulosic polysaccharides are synthesized by enzymes encoded by eight or ten multigene families of *cellulose synthase-like* (*Csl*) genes, which share sequence homology with the *CesA* genes. Most CSL proteins do not contain, however, the zinc-finger domain and/or other plant-specific sequences present in CesAs (Delmer 1999). The precise functions of the CSL families are under active investigation. The *CslF* and *CslH* genes appear to be restricted to the Poaceae, whereas the *CslB* and *CslG* genes are only present in eudicots (Fincher 2009; Penning et al. 2009). The close evolutionary relationships among the grasses are used to infer functions across species. With the use of Web-based databases, genome browsers, and mutant collections (see Sect. 5), it is possible to identify the closest sorghum orthologs for genes identified in other species. Below follows a brief summary of recent discoveries related to genes involved in the biosynthesis of hemicellulosic polysaccharides.

The *CslF* gene family encodes enzymes that catalyze the biosynthesis of mixed-linkage β-glucans based on observations in rice (Burton et al. 2006) and wheat (Nemeth et al. 2010). Sorghum appears to have retained more of the duplicate copies of these genes than maize, representing the opposite situation of what was observed for the *CesA* genes (Schnable et al. 2009; Sect. 3.1). Xyloglucan, a polysaccharide that is present in relatively small amounts in grass cell walls, is synthesized by members of the CSL-C family (Cocuron et al. 2007). *CslD* genes have been implicated in cellulose synthesis in the growing tips of maize roots (Kim et al. 2007). The *CslA* genes have been shown to be involved in the synthesis of β-mannans (Dhugga et al. 2004; Liepman et al. 2005), a storage polysaccharide present in grain. It was initially postulated that the xylan backbone of GAX would be synthesized by members of a CSL family, but analysis of gene expression data from a variety of species points towards certain glycosyltransferases (GT) as more likely candidates (Mitchell et al. 2007). GTs are encoded by large multigene families and are non-iterative enzymes, i.e., they tend to catalyze single glycosylation reactions such as the ones involving the addition of arabinose to the xylan backbone of GAX (Mitchell et al. 2007). The experimental proof for GT-mediated xylan biosynthesis is, however, still lacking.

After their initial biosynthesis and deposition in the cell wall, the hemicellulosic polysaccharides often undergo remodeling, mediated by enzymes named xyloglucan endo-β-transglucoslyase/hydrolases (XTHs) (Thompson and Fry 2001; Rose et al. 2002). These enzymes can break backbone structures and attach them to other cell wall polymers of the same or different origin.

Detailed descriptions of the different classes of enzymes involved in the biosynthesis of hemicellulosic polysaccharides are provided by Girke et al. (2004), Fincher (2009), Penning et al. (2009), and Faik (2010).

3.3 Monolignol Biosynthetic Genes

Whereas sorghum has played a relatively minor role in the elucidation of genes involved in polysaccharide synthesis, several genes involved in lignin biosynthesis were identified by Porter et al. (1978), who reported on a collection of chemically induced mutants with a *brown midrib* phenotype. The midrib is the main vascular structure in the center of the leaf blade, and, as the name implies, the *brown midrib* mutants are characterized by a reddish-brown coloration of the midrib. Upon closer examination it is evident, however, that all vascular bundles in the leaf blade and leaf sheath are brown. In some, but not all, of these mutants the vascular tissue in the stems and roots is brown as well. In sorghum the abbreviation *bmr* is used to designate the *brown midrib* mutants, whereas in maize a phenotypically similar class of mutants is designated as *bm* (reviewed by Vermerris (2009) and Sattler et al. (2010)). In sorghum the abbreviation *bm* had already been used to refer to *bloomless* mutants (Ayyangar and Ponnaiya 1941).

The collection generated by Porter et al. (1978) included 24 mutants which they numbered consecutively *bmr1* through *bmr24* without having established allelic relationships. A number of these mutants were lost as a result of either deleterious effects of the *bmr* mutations themselves, or the presence of additional mutations that reduced viability. An additional set of four spontaneous mutants was discovered later on, and more recently another large set of *bmr* mutants were identified in a TILLING population developed by Xin et al. (2008) (see also Sect. 5.2 and Chap. 8).

Allelic relationships of the original 24 chemically induced and four spontaneous *bmr* mutants were investigated and reported by Saballos et al. (2008). Four different allelic groups were identified based on a combination of crosses and chemical analyses and were represented by the mutants *bmr2*, *bmr6*, *bmr12*, and *bmr19*. These four mutants were designated as the reference (*-ref*) mutants, and a naming convention representative of the allelic relationships while retaining the original designator was proposed for the complete set of mutants. For example, *bmr3* was renamed *bmr6-3* to reflect the fact that *bmr3* and *bmr6* are allelic. Some of the *bmr* alleles reported by Xin et al. (2008) have been characterized and reported to be alleles of *bmr12*. Allelism tests and additional characterizations are under way.

The *Bmr12* gene was the first sorghum *Bmr* gene to be cloned. Based on the chemical changes observed in the lignin—a strong reduction in syringyl residues and an increase in guiacyl residues—the gene encoding caffeic acid *O*-methyltransferase (COMT) was proposed as the candidate gene. Based on its activity, a more accurate name of this enzyme is 5-hydroxyconiferylaldehyde/alcohol *O*-methyltransferase (Humphreys et al. 1999). Indeed, the *COMT* gene, present as a single copy, was

shown to contain mutations in the allelic mutants (prior to renaming) *bmr12*, *bmr18*, and *bmr26* (Bout and Vermerris 2003). In all three cases the mutations led to premature stop codons that resulted in strongly reduced levels of *COMT* mRNA. A similar approach was taken to clone the *Bmr6* gene. The *bmr6-ref* mutant contains increased levels of coniferaldehyde endgroups in its lignin, which is particularly clear after staining with the Wiesner reagent (acid phloroglucinol), as well as chemical analytical methods (Pillonel et al. 1991; Saballos et al. 2008; Sattler et al. 2009). The increase in cinnamaldehydes can be explained by a reduced activity of the enzyme cinnamyl alcohol dehydrogenase (CAD). Pillonel et al. (1991) indeed showed reduced CAD activity in *bmr6*. The cloning of the *Bmr6* gene was, however, complicated by the existence of a *CAD* multigene family. The *Bmr6* gene was independently cloned by Saballos et al. (2009) and Sattler et al. (2009). Saballos et al. (2009) performed a genome-wide scan of the sorghum genome sequence (Paterson et al. 2009) to identify all 14 *SbCAD* genes (including four pseudogenes), which were numbered using the rice genome as a reference. They then used a combination of mapping and sequencing of multiple independent *bmr6* alleles (*bmr6-ref*, *bmr6-3*, and *bmr6-27*) to prove that *Bmr6* encodes SbCAD2, a protein with strong similarity to known lignification-related CADs from several other species. Allele-specific markers for each of the mutant alleles were developed to facilitate breeding efforts using these mutations. Sattler et al. (2009) cloned and sequenced the sorghum orthologs of the Arabidopsis *AtCAD4* and *AtCAD5* genes, known to play a role in monolignol biosynthesis (Kim et al. 2004; Sibout et al. 2005), from a set of near-isogenic *bmr6* lines in three genetic backgrounds: "Atlas," "Wheatland," and RTx430. They demonstrated that in each of the *bmr6* mutant lines the *AtCAD5* ortholog contained a nonsense mutation. Activity assays showed that the enzyme encoded by the *Bmr6* gene displays higher activity towards sinapyl aldehyde (**8**; Fig. 17.2) than coniferyl aldehyde, which explains the decrease in the S:G ratio in the lignin of the *bmr6* mutant. Comparison of enzyme kinetics between the CAD protein encoded by *Bmr6* and the sorghum ortholog of *AtCAD4* (Sb02g024190) indicated that the latter plays only a minor role in the biosynthesis of monolignols.

The three near-isogenic *bmr6* lines used by Sattler et al. (2009) were part of a larger collection of near-isogenic *bmr6* and *bmr12* mutants developed by Pedersen et al. (2008) that also contained the grain sorghums "Redlan," BTx623, BTx630, and BTx631. Earlier efforts by the same group resulted in near-isogenic *bmr6* and *bmr12* mutants in the forage sorghums "Atlas," "Kansas Collier," "Rox Orange," and "Early Hegari Sart" (Pedersen et al. 2006). Comparing single and double mutant combinations in "Wheatland" and RTx430 backgrounds, Palmer et al. (2008) reported that the genetic background influenced the extent to which the *bmr* mutations affected both the amount of soluble phenolic compounds and cell wall composition, specifically lignin subunit composition and the amount of ester-bound hydroxycinnamic acids. These data suggest that by selecting specific genetic backgrounds, the impact of these *bmr* mutations can be enhanced or mitigated, which is of importance when considering commercial production of cultivars and hybrids harboring *bmr* mutations.

4 Breeding Targets to Enhance Sorghum Biomass Utilization

4.1 Forage and Silage Quality

Total dry matter intake and energy availability for ruminants are influenced by cell wall digestibility. Lignin is a ubiquitous component of the cell walls and lignin content has been shown to be negatively correlated to cell wall digestibility. Consequently, one of the most effective ways to improve digestibility is to reduce the lignin content (Cherney et al. 1991a). The *bmr* mutations (Sect. 3.3) have been a useful tool to modify lignin content and composition. Commercial forage sorghum and sudangrass containing the *bmr6-ref* and *bmr12-18* (*bmr18*) mutations are available and the presence of these mutations has been shown to reduce the ADL by 21 and 13 %, respectively, compared to their wild-type counterparts (Oliver et al. 2004). These changes translate in increases of 4 kg day^{-1} in milk production for dairy cows fed with *bmr* silage (Oliver et al. 2004) and 0.22 kg day^{-1} weight gain for steers that grazed residues from grain sorghum carrying the *bmr* trait (Schwarz et al. 2008). The effect of the *bmr* mutations on lignin content and in vitro and in vivo digestibility appears, however, to depend on the genetic background of the line (Oliver et al. 2005a, b) and the contribution of the *bmr* trait is not always positive when total dry matter yield per area of production is taken into consideration (Miron et al. 2007).

In addition to modifications in lignin content, modifications in plant architecture that increase the leaf-to-stem ratio can be beneficial. Leaf blade tissue contains less ADL and more hemicellulosic polysaccharides than stem tissue (Gerhardt et al. 1994), so that the fraction of easily digestible tissue in the total plant biomass is bigger. The higher digestibility decreases biomass retention time in the rumen and, therefore, increases the amount of material consumed (Cherney et al. 1991b). Sudangrass and sorghum × sudangrass hybrids have a higher leaf-to-stem ratio than grain and sweet sorghum and have traditionally been used for forage, hay, and silage production for this reason. An additional benefit of the higher leaf-to-stem ratio is that the resulting changes in grass canopy architecture improve ingestive behavior of grazing livestock (Chacon and Stobbs 1976; Burlison et al. 1991).

Plant anatomy at the cellular level has been shown to influence forage digestibility in bermudagrass (Ehlke and Casler 1985) and orchardgrass (Casler and Carpenter 1989). A lower proportion of highly lignified tissues such as sclerenchyma in the leaf blade area would be expected to result in increased digestibility, all other factors being equal. Although no systematic study of genetic variability with respect to the percentage of the different tissues in the leaf and stems of sorghum has been undertaken, natural variation is likely to exist and could be used to improve forage quality.

Under some environmental conditions, forage sorghum can accumulate compounds that are toxic to livestock (Pedersen and Fritz 2000). The cyanogenic glucoside dhurrin (**10**; Fig. 17.2) releases hydrogen cyanide (HCN) when degraded by enzymes in the plant or in the rumen (Wheeler and Mulcahy 1989). Considerable variation

between genotypes in HCN potential has been reported (Gorz et al. 1986, 1987; Lamb et al. (1987), but these studies did not agree on the inheritance of the trait. Gorz et al. (1986) identified a pair of codominant alleles controlling the trait, whereas a segregation study by Lamb (1987) indicated dhurrin concentration was a quantitative trait controlled by several loci with additive effects. Multigenic inheritance of the trait was reported by Gorz et al. (1987). This study involved chromosome translocation lines, and many chromosome fragments contributed significantly to dhurrin concentration. More recently, Mohanraj et al. (2006) concluded that the trait is controlled by nonadditive gene action based on the combining ability for dhurrin concentration of sorghum inbred lines. At the molecular level dhurrin concentration is determined by the transcriptional control of the genes encoding two cytochrome P450 enzymes, CYP79A1 and CYP71E1 (Busk and Møller 2002). Additional gene expression studies using microarrays or next-generation sequencing may reconcile the different findings.

Nitrate accumulation can also present problems of toxicity in livestock. Environmental stress such as drought or frost and high levels of nitrogen fertilization cause high nitrate accumulation in sorghum. The nitrate levels on forage sorghum are usually managed through agronomic practices and ensiling. Due to a lack of studies it is unclear to what extent genetic factors control this trait.

4.2 Cellulosic Ethanol

Cellulosic ethanol is produced from the microbial fermentation of monosaccharides generated via the enzymatic and/or chemical hydrolysis of cell wall polysaccharides present in plant biomass. The biomass is processed in a biorefinery (Ragauskas et al. 2006). First, the particle size of the biomass is reduced (if necessary), and then the biomass is subjected to a thermochemical pretreatment that generates soluble sugars from the hemicellulosic polysaccharides and enables access to the cellulose (Mosier et al. 2005). The cellulose is hydrolyzed to glucose by cellulolytic enzymes, typically of microbial origin. The glucose (and in some cases, the pentose sugars) are fermented to ethanol by either yeast or bacteria. Several processing schemes have been developed or proposed to enhance the efficiency of these different processing steps, including simultaneous saccharification and fermentation (SSF), whereby the biomass is hydrolyzed inside the fermentation tank, and consolidated bioprocessing (CBP), whereby the production of hydrolytic enzymes occurs concurrently with saccharification and fermentation (Lynd et al. 2005). Regardless of the exact process, the ethanol ultimately needs to be distilled from the beer.

The thermochemical pretreatment is necessary because biomass is inherently recalcitrant, i.e., resistant to degradation, in order to ensure the plant's survival up to and beyond seed production. The typical yield of glucose after extended enzymatic hydrolysis with cellulolytic enzymes is typically less than 25 % of the theoretical maximum (Mosier et al. 2005; Li et al. 2010). The recalcitrance is the result of the tightly packed and crystalline structure of cellulose, the heterogeneity of the

hemicellulosic polysaccharides and the presence of lignin in the cell wall. The cellulolytic enzymes can irreversibly adsorb to the lignin and lignin also provides a physical barrier that limits access to the cellulose (Palonen et al. 2004). Thermochemical pretreatments can raise the yield of fermentable sugars obtained after enzymatic saccharification to 90 % (Mosier et al. 2005), but this process is energy intensive and costly. Reductions in lignin content and changes in lignin subunit composition have been shown to enhance the yield of fermentable sugars in a number of species, including maize (Vermerris et al. 2007), alfalfa (Chen and Dixon 2007), and poplar (Davison et al. 2006).

Enhanced sugar yields were also observed when stover from sorghum with altered lignin composition was used (Vermerris et al. 2007; Saballos et al. 2008; Dien et al. 2009). The *bmr12* mutation resulted in a 1.25- and 1.75-fold increase in hexoses (mostly from cellulose) and pentoses (from hemicellulosic polysaccharides), respectively, in the hydrolysate obtained after enzymatic saccharification (Vermerris et al. 2007). Saballos et al. (2008) reported increases in the yield of glucose of 17, 20, 7, and 21 %, respectively, for *bmr2*, *bmr3* (=*bmr6-3*), *bmr6*, and *bmr12* using unpretreated stover, After dilute acid pretreatment, the overall yield of glucose greatly increased to an average of 90 % conversion, but the percent increase resulting from these *bmr* mutations remained identical. Consequently, these mutations can be viewed as a genetic pretreatment that is effective even in combination with thermochemical pretreatments. The data reported by Dien et al. (2009) on *bmr6*, *bmr12*, and *bmr6–bmr12* double mutants corroborate this observation. These authors reported that the effect of the two individual mutations acted in an additive manner in the double mutant, both in terms of Klason lignin content (−13, −15, and −27 %, respectively) and conversion of cellulose to ethanol after dilute acid pretreatment followed by SSF with *Saccharomyces cerevisiae* (+22, +21, and +43 %, respectively). Processing of stover from the double mutant also produced more ethanol after dilute alkali (ammonium hydroxide) pretreatment followed by SSF, although the improvement (+8 %) was not as striking as with the dilute acid pretreatment, possibly due to the fact that the processing conditions had not been optimized. Despite suggestions that Klason lignin content can adequately predict glucose or ethanol yields (Chen and Dixon 2007; Dien et al. 2009), the effect of lignin subunit composition or the presence of phenolic intermediates resulting from metabolic perturbations cannot be excluded. Enzymatic saccharification of *bmr19* stover reportedly led to lower glucose yields relative to the wild-type control, despite its lower Klason lignin content (Saballos et al. 2008).

An alternative pretreatment method is ammonia fiber expansion (AFEX) (Sendich et al. 2008), which involves exposing the biomass to ammonia under pressure (110–140 °C, 21 atm., 30 min). This process differs from most other pretreatment processes (Mosier et al. 2005) in that it returns a dry, activated biomass sample as opposed to a slurry. Li et al. (2010) optimized the conversion of biomass from forage sorghum and sweet sorghum bagasse using the AFEX-pretreatment. The cellulose and hemicellulosic polysaccharides from the pretreated biomass were hydrolyzed with cellullases and xylanases and the resulting monomeric sugars were fermented to ethanol by a transgenic yeast strain able to co-ferment hexoses and pentoses (Sedlak and Ho 2004). The processing of biomass from forage sorghum and sweet sorghum bagasse resulted in 90 and 97 % yield of

ethanol, respectively, as long as the bagasse was washed prior to the pretreatment to recuperate soluble sugars. These soluble sugars could be added into the reactor used for enzymatic hydrolysis so that they would be available for fermentation. The AFEX process resulted in the degradation of soluble sugars that remained in the sweet sorghum bagasse.

Screening of a sorghum diversity panel consisting of 381 entries with the use of a miniaturized assay for enzymatic saccharification of stover, with or without ammonium hydroxide pretreatment, demonstrated considerable genetic variation for rate of hydrolysis within the species (Vandenbrink et al. 2010). These results open the possibility of mapping genes associated with high yields of fermentable sugars and introgressing favorable alleles into commercial varieties.

In reviewing these data, the target for breeding of sorghums intended to be used for cellulosic ethanol production thus appear similar to the target of forage breeding programs, which is to increase the content of polysaccharides, decrease the content of lignin and/or modify lignin subunit composition, while at the same time increasing the overall harvestable biomass yield. One trait that does not need taken into to consideration when developing sorghums for biofuels is palatability.

4.3 Biomass for Combustion

While in the short-to-medium term sorghum biomass is most likely going to be used for the production of ethanol and other chemicals, it is technically possible to simply burn the biomass to generate heat (thermal energy). The heat can be used to distill starch- or sugar-based ethanol from the beer (see Sect. 4.2), or it can be used to produce the steam necessary to drive a turbine to generate electricity. In these cases, the objective is to maximize energy output, which can be achieved by maximizing the proportion of reduced (as opposed to oxidized) carbon in the biomass. Consequently, given the abundance of reduced carbon atoms (C–C and C–H linkages) in lignin, biomass with a high lignin-to-carbohydrate ratio is desirable. Given that sorghum biomass contains approximately 20 % lignin, whereas wood from hardwood trees contains as much as 40 % lignin, sorghum is not a very efficient source of heat. Its low bulk density, especially compared to wood, and its susceptibility to rot when stored over the winter, further limit its attractiveness. A viable option is to burn the lignin that is generated as a waste product from the biorefinery. This may be more appealing that processing the lignin into various low-value products, such as adhesives, for which the market will likely get saturated very quickly once biorefineries are operating on a large scale.

4.4 Drop-in Fuels

The term “drop-in fuels” refers to alternative fuels (as opposed to fossil fuels) other than ethanol that are generally more similar to or more compatible with current fossil fuels.

An example of a drop-in fuel is biodiesel, which is produced through the *trans*-esterification of fatty acids from plant and algal oils and lipids. Bio-diesel can be used in many conventional diesel engines.

The interest in developing drop-in fuels is driven by the desire to use fuels with a higher energy content than ethanol (on a weight or volume basis), so that they can be used in high-powered engines in airplanes, trucks, and trains. Drop-in fuels are currently not being produced on a large scale. Several synthesis routes have been proposed, including production of higher alcohols such as butanol and iso-butanol through microbial fermentation of monosaccharides from biomass (Ezeji et al. 2007), fuels derived from butanol and iso-butanol through chemical reactions, the production of pyrolysis oils generated by heating biomass under low- or no-oxygen conditions followed by synthesis of fuels with the help of chemical catalysts analogous to the production of gasoline, and the chemical synthesis of fuels from a syngas (Fischer-Tropsch fuels), obtained from heating biomass under low-oxygen conditions (Hamelinck and Faaij 2006).

Drop-in fuels are considered to represent the next generation of biofuels after cellulosic ethanol, and the biomass counterpart to biofuels derived from algae. The lack of a clear winner among the drop-in fuels and the different synthesis protocols make it challenging to formulate breeding targets. In the case of (*iso*-)butanol derived from fermentable sugars, high-biomass sorghums with a low lignin-to-polysaccharide ratio will be favored, whereas the pyrolysis and syngas routes will likely be more flexible in terms of biomass composition, as temperature and oxygen levels can be varied relatively easily.

5 Genetic and Genomic Approaches to Enhance Biomass Utilization

5.1 Natural Genetic Variation

Due to its extensive distribution range, sorghum includes a wide range of morphological and developmental types adapted to different growing conditions. Variation in plant height, maturity, and photoperiod is easily accessible within the primary gene pool. Genes for perenniality exist in many species within the genus Sorghum. The amount of genetic variation within the improved sorghum germplasm is small compared to the variation existing in the landraces and wild sorghums (Menz et al. 2004).

Germplasm of improved cultivars, landraces, and wild relatives of sorghum can be freely obtained from the Germplasm Resource Information Network (GRIN, http://www.ars-grin.gov) for genetic research and breeding programs. The International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) also hosts a large collection of cultivated and wild sorghum entries. (http://www.icrisat.org/sorghum/Project1/pfirst.asp). Because many of the entries from the same geographical areas are genetically similar, efforts have been made to define a subset

of entries that encompass most of the genetic spectrum of the species. Samples from the Sorghum World Collection have been selected to form core collections based on morphologic, agronomic, taxonomic, and geographic information, and more recently, also based on molecular marker data (Grenier et al. 2001a, b; Djè et al. 2000). China manages a Chinese Sorghum Germplasm diversity collection (http://icgr.caas.net.cn/ cgris_english.html). The majority of those entries are not present in the ICRISAT or U.S. collections.

The natural genetic variation that exists within the species can be exploited for the improvement of biomass traits. In addition to modifying cell wall composition (see Sect. 4.2), traits to be considered are plant architecture and biomass yield, each discussed in more detail below.

Plant architectural traits control the number and size of the leaves, the leaf area index (a function of the leaf angle), the diameter of the stem, the number of vascular bundles per stem, the number of tillers, and the size of the panicle. Maximum biomass yields will be obtained with plants that are tolerant to high plant density, which can be attained by having plants with erect leaves (high leaf area index) that are able to capture the incoming solar radiation effectively. A large stem diameter contributes to biomass yield per plant, but depending on physical characteristics of the stem, may lead to lodging due to the weight of the plant, or stalk breakage due to wind or becoming top heavy when the panicle emerges. The characteristics of the sclerenchyma (thickness, spacing) and the number of vascular bundles and their composition, will play an important role in determining the physical characteristics of the stem. As discussed for forage quality (Sect. 4.1), the relative proportion of the different plant organs may influence the digestibility of the biomass. Sudangrass and sorghum × sudangrass hybrids continue to produce profuse tillers throughout the growing season, with a high leaf-to-stem ratio. In addition, the stems of forage varieties are slender, which decreases the dry-down time of the biomass, facilitating the harvest and storage.

Height and maturity in sorghum are controlled by the *Dwarf* (*Dw*) and *Maturity* (*Ma*) genes, respectively. The role of these genes is discussed in detail by Brown and Paterson in Chap. 14 of this book. Grain sorghums benefit from reduced height because of their compatibility with combine harvesting methods, as well as the reduced susceptibility to lodging due to the presence of a large panicle. In contrast, forage and bioenergy production benefits from the increased height associated with wild-type *Dw* alleles.

Biomass yield is a function of plant height, the number of stems and dry matter content. The highest biomass yields will generally be obtained from plants that produce tall, thick stems, and many tillers, and that can be grown at relatively high plant density. Biomass yields of sorghums specifically grown for this purpose average 15–25 Mg dry matter (DM) ha^{-1}, but may be as high as 40 Mg DM ha^{-1} (Rooney et al. 2007; Venuto and Kindiger 2008) depending on the location and genotype. High biomass yields can be obtained by using photoperiod-sensitive or late-maturity varieties that have a long vegetative phase, which allows the plants to continually produce internodes and therefore accumulate biomass during the entire growing season (Rooney et al. 2007). However, considering biomass yield per unit land area

and per growing season (as opposed to per harvest), maximum biomass yields may be achieved by planting faster growing plants that mature more quickly. In this case, a second harvest from a ratoon crop or a second planting (from new seed) may exceed the biomass yield obtained from a single harvest of late maturity types. This may be especially true when sudangrass or sorghum × sudangrass hybrids are selected, because they can have perennial or semi-perennial growing behavior. In order to investigate this in detail, Venuto and Kindiger (2008) compared biomass yields of one sudangrass cultivar, ten sorghum × sudangrass hybrids, and ten hybrid forage sorghums including late-maturity photoperiod-sensitive hybrids, grown in replicated field trials in Oklahoma. They compared biomass yield from a single harvest with that from a double harvest (main harvest and ratoon). On average, the highest yields were obtained after a single harvest, with a maximum reported DM yield of 40.3 Mg ha^{-1} from the hybrid forage sorghum "Tentaka." On an individual basis only a few entries produced higher biomass yields from two harvests. Tew et al. (2008) reported superior biomass yields of the nonflowering sorghum × sudangrass hybrid MMR 333/47 compared to several sweet sorghum varieties when grown over a long season without ratooning. This study was aimed at determining the best crop to grow in rotation with sugarcane in Louisiana, with (theoretical) ethanol yield—either from biomass alone or from biomass and soluble sugars combined—being the main driver (see Sect. 4.2). Given the 120-day window in which the sorghum had to be harvested, processing of the hybrid MMR 333/47 or the sweet sorghum "Theis" would result in the largest volume of ethanol.

It is important to consider that the reported biomass yields originate from studies at specific geographic locations, and that differences in temperature, photoperiod, soil, planting date, and genotype all contribute to differences in biomass yields. Consequently, in order to determine the most effective production system for a given location, it would be best to conduct a study at that location with the most advanced germplasm available.

5.2 Chemically Induced Mutants

In the absence of efficient and reliable transposable element systems such as in maize (McCarty et al. 2005; Bai et al. 2007), a T-DNA tagging system such as in Arabidopsis (Alonso et al. 2003), and in light of the risk of transgenes escaping because of the facile cross-fertilization with the related weedy species Johnsongrass (*Sorghum halepense* (L.) Pers.), the only convenient and efficient alternative to induce genetic variation in sorghum has been mutagenesis. The first reported attempt to generate chemically induced mutants of sorghum was intended to generate *bloomless* (*bm*) mutants and also resulted in a large number of *brown midrib* (*bmr*) mutants (Porter et al. 1978; Sect. 3.3). More recently, a TILLING population was developed by Xin et al. (2008). TILLING stands for Targeted Induced Local Lesions IN Genomes and involves the generation of mutants resulting from chemical mutagenesis with the

alkylating agents ethyl methane sulfonate (EMS) or diethylsulfate (DES) followed by the cataloguing of both visible mutant phenotypes (forward genetics) and induced DNA polymorphisms (reverse genetics). A detailed description of this particular resource is provided by Xin et al. in Chap. 8 of this book. This TILLING population contains several mutants that are of interest for bioenergy applications, including novel *bmr* mutants and plants with modified architecture (Xin et al. 2009).

5.3 Genomics Approaches to Identify Cell Wall-Related Genes

With the availability of whole-genome sequences for many different plant species, it has become feasible to identify target genes in sorghum based on a combination of bioinformatics, genomics, and experimental data in other species. The use of the candidate-gene approach to identify and clone *Brown midrib* genes as described in Sect. 3.3 is an example of the use of genomics resources in such a manner. In these instances the *Bmr* genes were identified based on phenotypic data obtained from existing mutants. An alternative to the use of mutants is the identification of target genes based on other types of data, such as QTL or gene expression data that imply the importance of certain genes or genomic fragments in a process of interest, or the availability of transgenics in which the expression of a gene of interest has been modified.

QTL mapping studies for forage and cell wall characteristics in maize can be directly translated to sorghum because of the close evolutionary relationship of these two species. For example, Ralph et al. (2004b) summarized the information on locations involved in cell-wall lignification or digestibility identified in the maize genome. Candidate genes involved in lignin biosynthesis are located in several of these regions, giving support to their role in cell wall digestibility traits. Several cell wall digestibility QTLs that are not colocalized with QTL related to lignin have also been identified, indicating that other cell wall or plant morphology-related genes influence digestibility (Riboulet et al. 2008). The homeologous regions of the sorghum genome are likely to contain genes of interest for these traits.

As a model plant for which a vast amount of information is available, Arabidopsis represents a good source of candidate genes for cell wall and plant architecture traits. However, due to the evolutionary distance between the two species caution should be exerted when postulating putative functions for sorghum homologs. As exemplified by an evolutionary study of lignin biosynthesis genes (Xu et al. 2009), genes of grasses have undergone rapid evolution since the split of monocots and dicots. This functional divergence is what likely led to the different cell wall architecture between the two groups.

While the identification of homologs of structural genes may be fairly straightforward, it is much more challenging to identify homologs of regulatory genes, in part because the presence of conserved DNA-binding domains may not imply conservation of target genes. Shen et al. (2009) recently identified regulatory genes for cell wall

biosynthesis using a comparative genomics approach. NAC proteins are part of a large family of plant-specific transcription factors, with hundreds of members in each of the species that have been investigated (Ooka et al. 2003; Olsen et al. 2005). Arabidopsis *NAC* genes have been implicated in developmental processes, stress response, and cell wall biogenesis. Shen et al. (2009) exploited phylogenetic relationships of *NAC* genes from 11 organisms including sorghum, and combined this information with the data from the characterized Arabidopsis *NAC* genes to classify all *NAC* genes into groups with distinct functions. Those genes likely involved in the regulation of cell wall biogenesis were further analyzed through motif grouping combined with co-expression data of *NAC* and cell wall biosynthesis genes based on publicly available microarray data. This approach was validated by the identification of a mutant of alfalfa (*Medicago truncatula*) that contained a retrotransposon insertion in the *NAC* gene *Secondary wall thickening promoting factor1* (*MtNST1*; Zhao et al. 2010). The expression of several lignin biosynthetic genes was reduced in this mutant, which resulted in a lower lignin content and altered patterns of lignin autofluorescence.

Several useful databases have been developed that contain information on cell wall-related genes and that are excellent resources when considering a comparative genomics approach to identify target genes. These resources have been reviewed by Vermerris (2009) and Cao et al. (2010) and include several databases relevant to researchers interested in the Saccharineae. The Cell Wall Genomics resource at Purdue University (http://cellwall.genomics.purdue.edu) is based on Arabidopsis, maize, and rice cell wall gene families (Penning et al. 2009). This database includes phylogenetic trees of gene families representing six different stages of cell wall biogenesis, as well as information on maize and Arabidopsis cell wall mutants. The Cell Wall Navigator resource (http://bioweb.ucr.edu/Cellwall) uses the same classification of cell wall biogenesis as the Cell Wall Genomics Web site and provides information on cell wall families involved in primary cell wall metabolism in Arabidopsis and rice based on genome sequences and protein and EST databases (Girke et al. 2004). This resource enables the exploration of evolutionary relationships within and between families of cell wall-related genes. The MAIZEWALL resource (http://www.polebio.scsv.ups-tlse.fr/MAIZEWALL) lists 735 maize cell wall-related genes identified based on homology and keyword searches of the GénoPlante-Info database (Samson et al. 2003) and GenBank (http://www.ncbi.nlm.nih.gov). In addition, expression data on 651 of these genes based on a "macroarray" are provided (Guillaumie et al. 2007). Combined with access to the sorghum genome sequence (http://www.phytozome.net/sorghum; Paterson et al. 2009) and Internet resources such as Gramene (http://www.gramene.org) and the Plant Genome Database (http://www.PlantGDB.org), these resources enable the identification of cell wall-related genes in sorghum and sugar cane.

6 High-Throughput Screening Methods to Identify Modified Cell Wall Composition

When the goal is to improve biomass utilization, it will be apparent from the previous section that there is no single breeding target. Instead,, the target depends largely on the specific anticipated end-use of the biomass. The best documented approach to improve biomass conversion has been the breeding for enhanced digestibility of forage and silage sorghums, which has resulted in commercial successes. Selection for cell wall composition or biomass conversion can be based on a variety of screening protocol that vary from each other in their specificity, equipment cost, and throughput. The most promising methods are described below.

6.1 Forage Quality Analysis

The traditional analysis of forage quality through the determination of NDF, ADF and ADL (see Sect. 2) is often considered too labor intensive for large numbers of samples. Ankom (Fairview, NY; www.ankom.com) has developed a semi-automated process for the analysis of forage quality using an incubator in which multiple small sample bags containing forage samples can be placed. Solutions are sequentially added and removed. An alternative method relies on the use of rumen fluid from fistulated cows to determine the digestibility of the biomass. This method is referred to as the in vitro ruminal (IVR) digestion test. When using this method, a convenient way to assay the digestibility of the specimen is to monitor the increase in pressure in the head space of the vial (Theodorou et al. 1994). With the development of near-infrared spectroscopy combined with multivariate statistical analyses, large numbers of samples can be processed. This is described in more detail in the next section.

6.2 Near-Infrared Reflectance Spectroscopy

Near-infrared reflectance spectroscopy (NIRS) is a vibrational spectroscopic technique in which the amount light in the near-infrared range of the electromagnetic spectrum (800–2,500 nm) that is reflected by a sample is measured and used to infer its chemical composition. The near-infrared light is absorbed by molecular bonds undergoing bend or stretch vibrations that alter the dipole moment (Siesler et al. 2002). Different functional groups (e.g., -OH, -NH_2, -COOH) and different types of vibrations will absorb light of different wavelengths. The energy of light in the mid-infrared range of the spectrum (2,500–4,000 nm) is associated with fundamental vibrational transitions, i.e., quantum transitions from one vibrational state to the next. The higher energy of light in the near-infrared region of the spectrum is associated

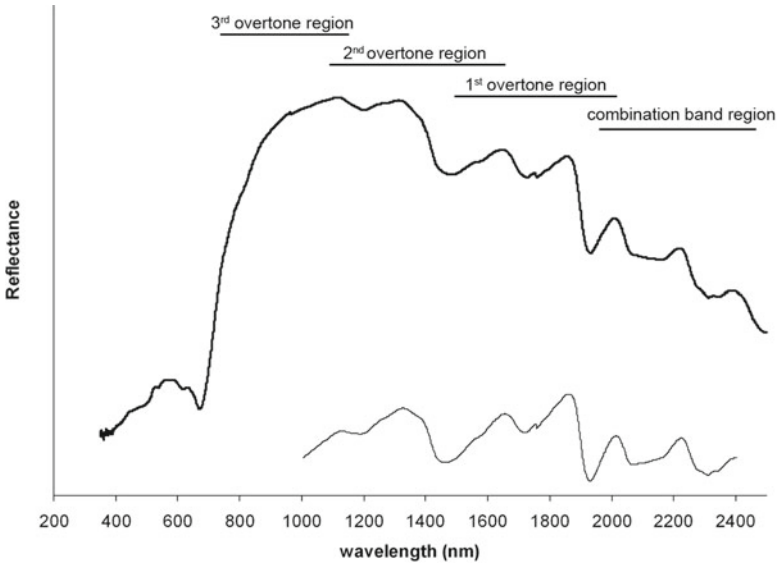

Fig. 17.4 Visible and near-infrared reflectance spectrum of a dried mature sorghum leaf collected from a wild-type BTx623 plant (*thick black line*). The horizontal axis represents the wavelength (350–2,500 nm), the vertical axis the reflectance (scaled to a maximum of 100 %). Multivariate analyses to identify putative mutants are performed on baseline-corrected and area-normalized spectra in the 1,100–2,400 nm range (*thin line*). The overtone and combination band regions are indicated

with overtones or combination vibrations, i.e., quantum transitions to other levels than the adjacent level.

Unlike many other chemical analytical techniques, NIRS is a nondestructive technique. NIR spectra are acquired with an NIR spectrometer that contains a polychromatic light source and a series of detectors recording the intensity of the reflected light at different wavelengths. Even though the acquisition of NIR spectra is quick and straightforward, an individual NIR spectrum (Fig. 17.4) is not very informative and it is typically difficult to deduce the exact composition of a sample based on a single spectrum. Data analysis generally requires the use of calibrations that rely on multivariate statistical models to correlate spectral data with compositional data obtained via more traditional analytical methods (cf. Jung et al. 1998; Brinkmann et al. 2002; Murray et al. 2008; Robinson and Mansfield 2009).

NIRS has also been used as a screening tool for the identification of novel maize mutants with altered chemical composition (Yong et al. 2005; Vermerris et al. 2007). Leaf segments were collected from individual plants in segregating F_2 families of the Uniform*Mu* population (McCarty et al. 2005). A collection of 39 mutants that, unlike the *brown midrib* mutants, had no apparent visible phenotypes under normal field conditions were identified (Penning et al. 2009).

We applied a similar approach to a subset of the sorghum TILLING population generated by Xin et al. (2008). Segments representing the center 10 cm of the fourth

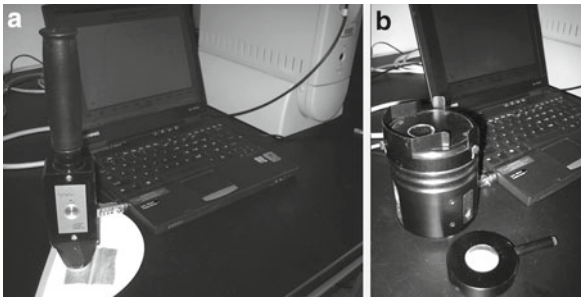

Fig. 17.5 (**a**) NIR spectrometer (FieldSpec Pro, Analytical Spectral Devices, Boulder, CO), equipped with a contact probe to acquire spectra from dried sorghum leaves sandwiched in between a GoreTex™ membrane and a piece of quartz glass. (**b**) The sample holder with bottom made of quartz glass is placed on top of a "mug light" to acquire spectra from powder

mature leaf were collected from approximately 200 field-grown segregating M_3 families in Lubbock, TX. The leaves were dried flat in envelopes to remove moisture and prevent microbial growth. NIR spectra were acquired for each of the leaf segments using a FieldSpec Pro NIR spectrometer (Analytical Spectral Devices, Inc.; Boulder, CO) equipped with a handheld contact probe. A section of the leaf blade was sandwiched in between a GoreTex™ membrane that was used as white reference (100 % reflectance) and a piece of quartz glass (Fig. 17.5a). The spectrum of each individual sample was based on the average of 30 spectra acquired in the 350–2,500 nm range, at 1-nm intervals. Area-normalized and baseline-corrected spectra (Fig. 17.4) from a given M_3 family were compared to spectra obtained from leaves of the BTx623 wild-type control, using a CLASS model in the WinDAS software package (Kemsley 1998). This unsupervised approach delineates the (multivariate) space occupied by wild-type samples, taking into account environmental variation in its broadest sense: plant-to-plant variation in the field, variation in sample collection (position of the leaf segment, time of day, leaf exposed to the sun, etc.), and variation during spectrum acquisition (leaf surface area smooth or wrinkled, pressure of the contact probe on the quartz glass, etc.). The CLASS model was based on a combination of principal components that together captured between 90 and 95 % of the variance among the samples in the wild-type group. Putative mutants were identified based on their position outside the wild-type space (Mahalanobis distance), calculated with a chi-square test. An additional constraint in evaluating the presence of mutants with spectral phenotypes is that the number of outliers has to be consistent with a Mendelian segregation pattern for a recessive or dominant mutation (Vermerris et al. 2007). A recessive mutation is expected to result in one-quarter of the plants displaying spectral characteristics different from the wild-type population, but similar to each other. The exact number of mutants per row varies stochastically, but the probability that a sample of 15 plants contains at least two mutants can be calculated using the binomial distribution and equals 0.920. When the number of

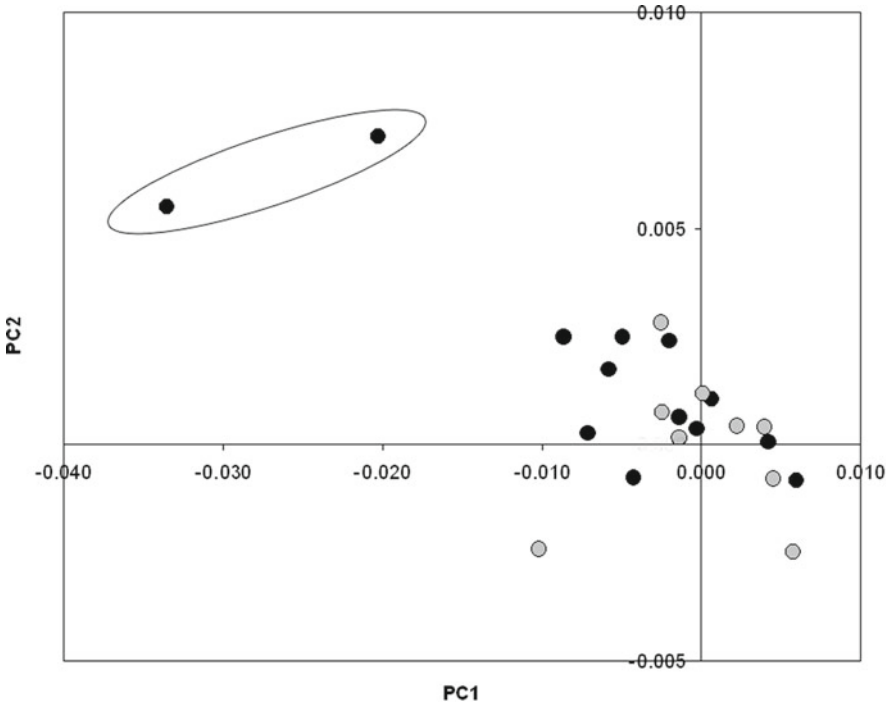

Fig. 17.6 Principal component score plot displaying spectral outliers in a segregating M_3 family. Each *circle* represents one individual plant. The wild-type control samples are represented by *gray circles*, whereas the M_3 family under investigation is represented by *black circles*. The *black circles* in the *ellipse* are spectral outliers based on CLASS modeling and represent putative mutants

plants per family is increased to 20, the probability of having at least two mutants increases to 0.976.

Figure 17.6 shows a principal component score plot to illustrate the identification of putative spectral mutants of sorghum. The gray symbols represent the wild-type control samples, and the black symbols the 13 leaf samples from a segregating M_3 family. The majority of these individuals fall within the wild-type space, but the two samples in the ellipse are outliers that cluster together. Enzymatic saccharification of stem tissue collected from these outliers can then be used to determine yields of glucose relative to the wild-type control. A convenient enzymatic saccharification assay is based on a protocol developed at the National Renewable Energy Laboratory (NREL) in Golden, CO (Brown and Torget 1996), with modifications described by Vermerris et al. (2007).

Based on different screens for NIR mutants we have conducted, up to 10% of the families in a mutant population contain spectral outliers suggestive of mutations. A subset of these families indeed contained plants with altered cell wall composition that led to more efficient biomass conversion. These results suggest that the NIRS-based screening of mutants to identify mutants with altered biomass conversion

properties is a promising strategy, especially since the majority of the families with outliers looked identical to wild-type plants, at least under standard field conditions. Re-screening of the progeny from a putative mutant is necessary to confirm that the spectral and biochemical changes are indeed the result of a genetic mutation.

NIR spectroscopy can also be applied to the analysis of ground samples, such as stover. This application is most common for forage quality analyses (see Sect. 6.1), especially when there is a good calibration set based on traditional forage quality analyses. The calibration will enable the prediction of composition or forage quality of an "unknown" sorghum sample based on its NIR spectrum (Murray et al. 2008; Venuto and Kindiger 2008). This approach has also been applied to the prediction of theoretical ethanol yield of biomass samples of maize (Lorenzana et al. 2010) and switchgrass (Mann et al. 2009).

For these types of analyses, a different sample holder is needed than the contact probe used with leaves. The fiber-optic cable of the FieldSpec Pro spectrometer can be inserted in a so-called "mug light," a canister with a light source inside and a window on top. The powder is then placed in a sample holder with a quartz glass bottom that is placed on top of the mug light (Fig. 17.5b). Similar configurations exist with other manufacturers. Spectral analysis of powders is challenging because the light tends to get scattered by particles on the surface, which reduces the signal-to-noise ratio in the spectra. In order to minimize sample-to-sample differences unrelated to the chemical composition, all samples need to be prepared in a consistent manner, i.e., ground through the same size screen, with the same kind of mill, and dried to the same moisture level.

NIR spectroscopy is typically used for the analysis of mature plant samples, such as forage or biomass for bioenergy, in part because of its sensitivity to aromatic compounds (lignin, hydroxycinnamic acids), and in part because of its ability to penetrate the sample beyond the surface layer. If the interest is in the identification of differences among genotypes in primary cell wall composition and architecture, however, it is possible and preferable to use Fourier-transform infrared spectroscopy. This technique relies on the absorbance of light in the mid-infrared range (2,500–4,000 nm). The spectra tend to be more informative because of energy levels associated with fundamental vibrational transitions. It is therefore often possible to assign peaks to actual compounds. Since mid-infrared light does not penetrate as deeply into the sample, it is generally necessary to acquire spectra using isolated cell walls as opposed to intact tissues, or samples pressed in a thin disk of potassium bromide (KBr). Corredor et al. (2008) used FTIR spectroscopy to monitor changes in the chemical composition of sorghum biomass after pretreatment and enzymatic saccharification.

6.3 Analytical Pyrolysis

Analytical pyrolysis is a method to obtain detailed compositional analysis on complete cell wall samples, such as biomass or wood. During analytical pyrolysis of cell wall

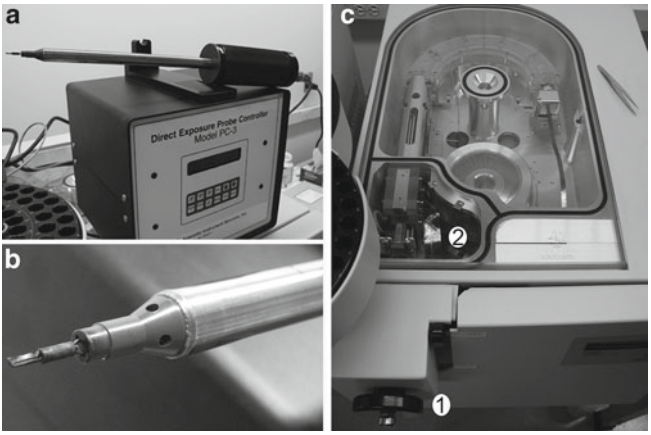

Fig. 17.7 (**a**) Ground tissue is placed on a resistively heated filament at the tip of a direct exposure probe. (**b**) Close-up of the filament. (**c**) The probe is inserted in a quadrupole mass spectrometer (shown from the *top*) through *inlet* **1** so that the tip is located inside ion source **2**

samples a volatile pyrolysate is generated as a result of thermal degradation under anoxic conditions (Evans and Milne 1987; Boon 1989). The pyrolysate contains breakdown products from the major cell wall constituents that can then be identified and quantified through either mass spectrometry (Py-MS) or gas chromatography–mass spectrometry (Py-GC–MS) (Boon 1989; Ralph and Hatfield 1991; Fontaine et al. 2003; Fig. 17.7). Py-MS is more amenable to high-throughput methods than Py-GC–MS, as analysis of a typical sample requires only a few minutes, whereas Py-GC–MS analyses require approximately 1 h per sample. The advantage of Py-GC–MS over Py-MS is its ability to resolve different fragment ions with the same mass-to-charge (m/z) ratio. Py-MS data can, however, be easily subjected to multivariate statistical analyses to identify the chemical basis for the classification of outliers that may represent putative cell wall mutants or transgenics in which cell wall composition has been modified.

Tar formed during pyrolysis is rather harsh on the components associated with the ion source of the mass spectrometer. Especially when pyrolysis is used for screening purposes, it is important to subject the equipment to maintenance on a regular basis in order to warrant consistency in the data.

High-throughput Py-MS screens to identify naturally occurring and transgenically induced variation in the composition of poplar and pine wood have been performed successfully (Davison et al. 2006; Sewell et al. 2002). No such screens have yet been reported in the literature for sorghum, but there is no reason why this approach would not work for the screening of sorghum populations. A major advantage of Py-MS over NIRS (Sect. 6.2) is that it is much more informative: it is possible to identify specific chemical differences as opposed to attribute differences to functional groups that may be shared among different cell wall constituents. This advantage of Py-MS may be outweighed, however, by the ease of use of NIRS especially when there is a good calibration model.

6.4 High-Throughput Biomass Conversion Assays

NIRS and Py-MS are indirect methods to predict biomass conversion efficiency or forage quality. Several more direct methods have been developed to assay biomass conversion assays.

A 96-well format pretreatment assay was developed by Studer et al. (2010). The method relies on small tissue samples that are loaded in a stainless steel 96-well plate, which can be placed in a pretreatment reactor where the samples can be subjected to high temperature and pressure. After the pretreatment is complete, the biomass can be subjected to enzymatic saccharification in the same reactor. Vandenbrink et al. (2010) developed a 96-well enzymatic hydrolysis assay to test the saccharification potential of a sorghum diversity panel consisting of 381 lines. The procedure was adapted from protocols developed at the National Renewable Energy Laboratory in Golden, CO (Dowe and McMillan 2008) and was performed with or without ammonium hydroxide pretreatment (15 % (w/v); 16 h, 65 °C). The concentration of reducing sugars released during the enzymatic saccharification was quantified using the colorimetric dinitrosalicylic acid assay (Miller 1959). While this assay is convenient, the sugar yield can conceivably be influenced by variation in soluble sugars and starch present in the biomass, especially in combination with the sterilization step (autoclave, 35 min, 121 °C) prior to hydrolysis of native and ammonium-hydroxide pretreated biomass.

Weimer et al. (2005) modified the in vitro ruminal assay described by Theodorou et al. (1994) (see Sect. 6.1) to assay biomass conversion instead of forage quality. A benefit of the IVR assay is that it is not necessary to sterilize the biomass. In addition, plant breeders and agronomists who traditionally focused on forage quality but are now considering bioenergy crops are familiar with the IVR assay. Biomass samples from switchgrass, eastern gamagrass and big bluestem samples were tested. Even though CO_2 production during fermentation with yeast was shown to be highly correlated to ethanol production, the correlation between gas production from IVR digestion of these same samples and ethanol yield resulting from fermentation appeared to vary with the origin of the sample (species). As long as ruminal fluid can be obtained, this method may be of interest for use as a primary screen for certain types of samples, but the lack of a strong correlation still requires the need for actual fermentation assays.

Haney et al. (2008) developed a fluorescence-based assay to measure sugar yields obtained from biomass. The assay relies on the growth of an *E. coli* strain that can utilize hexoses and pentoses as carbon source and that expresses the green fluorescent protein (GFP). When biomass obtained from near-isogenic maize hybrids with different *brown midrib* (*bm*) mutations was subjected to dilute acid hydrolysis and then simultaneous saccharification and catabolism, the higher biomass conversion efficiency of the *bm* samples was evident based on the higher level of fluorescence.

7 Conclusions

The low-input requirements, drought tolerance, and high biomass yield potential of sorghum make this species uniquely suited as a bioenergy crop. Since the efficiency of biomass conversion to fermentable sugars depends to a large extent on cell wall composition and plant architecture, the methodologies used by the traditional forage sorghum breeding programs can be implemented for the development of biomass sorghums, although modifications to some of the assays may be necessary. The genetic diversity of sorghum and the abundance of genomics resources make it feasible to expand the production area of sorghum beyond its current base. As long as uncertainty about the optimal biomass conversion process remains, a focus on low input requirements and high biomass yields should probably take precedence over specific compositional targets.

Acknowledgements The authors gratefully acknowledge support from the Office of Science (BER), U.S. Department of Energy, grant DE-FG02-07ER64458 for the research on *brown midrib* genes and NIRS screening of sorghum mutants described in this chapter. We would also like to acknowledge the participation of several colleagues on this project: The sorghum leaf samples were collected by Dr. Ken Lamb (University of Florida) in collaboration with Dr. Zhanguo Xin (USDA-ARS, Lubbock, TX) and his research staff. We thank Drs. Bryan Penning and Nick Carpita (Purdue University) for sharing their most recent data on maize and sorghum *CesA* genes. The U.S. National Science Foundation Plant Genome Research Program (DBI-0217552) and the University of Florida provided funds to purchase the analytical equipment featured in this chapter.

References

Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H et al (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science 301:653–657

Appenzeller L, Doblin M, Barreiro R, Wang H, Niu X, Kollipara K, Carrigan L, Tomes D, Chapman M, Dhugga KS (2004) Cellulose synthesis in maize: isolation and expression analysis of the cellulose synthase (*CesA*) gene family. Cellulose 11:287–299

Argillier O, Barrière Y, Hébert Y (1995) Genetic variation and selection criteria for digestibility traits of forage maize. Euphytica 82:175–184

Arioli T, Peng L, Betzner AS, Burn J, Wittke W, Herth W, Camilleri C, Höfte H, Plazinski J, Birch R, Cork A, Glover J, Redmond J, Williamson RE (1998) Molecular analysis of cellulose biosynthesis in *Arabidopsis*. Science 279:717–720

Ayyangar GNR, Ponnaiya BWX (1941) The occurrence and inheritance of a bloomless sorghum. Curr Sci 10:408–409

Bai L, Singh M, Pitt L, Sweeney M, Brutnell TP (2007) Generating novel allelic variation through *Activator* insertional mutagenesis in maize. Genetics 175:981–992

Barrière Y, Guillet C, Goffner D, Pichon M (2003) Genetic variation and breeding strategies for improved cell wall digestibility in annual forage crops. A review. Anim Res 52:193–228

Boon JJ (1989) An introduction to pyrolysis mass spectrometry of lignocellulosic material: case studies of barley straw, corn stem and *Agropyron*. In: Chesson A, Ørskov ER (eds) Physico-chemical characterization of plant residues for industrial and feed use. Elsevier Applied Science, London, pp 25–49

Bout S, Vermerris W (2003) A candidate gene-approach to clone the sorghum *Brown midrib* gene encoding caffeic acid *O*-methyltransferase. Mol Genet Genomics 269:205–214

Brinkmann K, Blaschke L, Polle A (2002) Comparison of different methods for lignin determination as a basis for calibration of near-infrared reflectance spectroscopy and implications of lignoproteins. J Chem Ecol 28:2483–2501

Brown L, Torget R (1996) Laboratory Analytical Protocol 009: enzymatic saccharification of lignocellulosic biomass. National Renewable Energy Laboratory, Golden, CO. http://www1.eere.energy.gov/analytical_procedures.html

Burlison AJ, Hodgson J, Illius AW (1991) Sward canopy structure and the bite dimensions and bite weight of grazing sheep. Grass Forage Sci 46:29–38

Burton RA, Wilson SM, Hrmova M, Harvey AJ, Shirley NJ, Medhurst A, Stone BA, Newbigin EJ, Bacic A, Fincher GB (2006) Cellulose synthase-like CslF genes mediate the synthesis of cell wall (1,3;1,4)-β-D-Glucans. Science 311:1940–1942

Busk PK, Møller BL (2002) Dhurrin synthesis in sorghum is regulated at the transcriptional level and induced by nitrogen fertilization in older plants. Plant Physiol 129:1222–1231

Cao P, Jung K-H, Ronald PC (2010) A survey of databases for analysis of plant cell wall-related enzymes. Bioenergy Res 3:108–114

Carpita NC (1996) Structure and biogenesis of the cell walls of grasses. Annu Rev Plant Physiol Plant Mol Biol 47:445–476

Carpita NC, Gibeaut DM (1993) Structural models of the primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the wall during growth. Plant J 3:1–30

Carpita NC, McCann MC (2008) Maize and sorghum: genetic resources for bioenergy grasses. Trends Plant Sci 11:314–320

Casler MD, Carpenter JA (1989) Morphological and chemical responses to selection for in vitro dry matter digestibility in smooth bromegrass. Crop Sci 29:924–928

Chacon EA, Stobbs TH (1976) Influence of progressive defoliation of a grass sward in the eating behaviour of cattle. Aust J Agric Res 27:709–727

Chen F, Dixon RA (2007) Lignin modification improves fermentable sugar yields for biofuel production. Nat Biotechnol 25:759–761

Cherney JH, Cherney DJR, Akin DE, Axtell JD (1991a) Potential of brown-midrib, low-lignin mutants for improving forage quality. Adv Agron 46:157–198

Cherney DJ, Mertens DR, Moore JE (1991b) Fluid and particulate retention times in sheep as influenced by intake level and forage morphological composition. J Anim Sci 69:413–422

Cocuron JC, Lerouxel O, Drakakai G, Alonso AP, Liepman AH, Keegstra K, Raikhel N, Wilkerson CG (2007) A gene from the cellulose *synthase-like C* family encodes a β-1,4 glucan synthase. Proc Natl Acad Sci U S A 104:8550–8555

Corredor DY, Salazar JM, Hohn KL, Bean S, Bean B, Wang D (2008) Evaluation and characterization of forage sorghum as feedstock for fermentable sugar production. Appl Biochem Biotechnol 158:164–179

Dahlberg JA (2000) Classification and characterization of sorghum. In: Smith CW, Frederiksen RA (eds) Sorghum. Origin, history, technology, and production. Wiley, New York, pp 99–130

Davison BH, Drescher SR, Tuskan GA, Davis MF, Nghiem NP (2006) Variation of S/G ratio and lignin content in a *Populus* family influences the release of xylose by dilute acid hydrolysis. Appl Biochem Biotechnol 129–132:427–435

De Boever JL, Cottyn BG, Andries JI, Buysse FX, Vanacker JM (1988) The use of cellulase technique to predict digestibility, metabolizable and net energy of forage. Anim Feed Sci Technol 19:247–260

Delmer DP (1999) Cellulose biosynthesis: Exciting times for a difficult field of study. Annu Rev Plant Physiol Plant Mol Biol 50:245–276

Desprez T, Vernhettes S, Fagard M, Refregier G, Desnos T, Aletti E, Py N, Pelletier S, Höfte H (2002) Resistance against herbicide isoxaben and cellulose deficiency caused by distinct mutations in same cellulose synthase isoform CESA6. Plant Physiol 128:482–490

Desprez T, Juraniec M, Crowell EF, Jouy H, Pochylova Z, Parcy F, Höfte H, Gonneau M, Vernhettes S (2007) Organization of cellulose synthase complexes involved in primary cell wall synthesis in *Arabidopsis thaliana*. Proc Natl Acad Sci U S A 104:15572–15577

Dhugga KS, Barreiro R, Whitten B, Stecca K, Hazebroek J, Randhawa GS, Dolan M, Kinney AJ, Tomes D, Nichols S, Anderson P (2004) Guar seed β-mannan synthase is a member of the *cellulose synthase* super gene family. Science 303:363–366

Dien BS, Sarath G, Pedersen JF, Satler SE, Chen H, Funnell-Harris DL, Nichols NN, Cotta MA (2009) Improved sugar conversion and ethanol yield for forage sorghum (*Sorghum bicolor* (L.) Moench) lines with reduced lignin contents. BioEnergy Res 2:153–164

Ding SY, Himmel ME (2006) The maize primary cell wall microfibril: a new model derived from direct visualization. J Agric Food Chem 54:597–606

Djè Y, Heuertz M, Lefèbvre C, Vekemans X (2000) Assessment of genetic diversity within and among germplasm accessions in cultivated sorghum using microsatellite markers. Theor Appl Genet 100:918–925

Dowe N, McMillan J (2008) SSF Experimental Protocols - Lignocellulosic Biomass Hydrolysis and Fermentation. Laboratory Analytical Procedure (LAP). Technical Report NREL/TP-510-42630. http://www.nrel.gov/biomass/pdfs/42630.pdf

Ehlke NJ, Casler MD (1985) Anatomical characteristics of smooth bromegrass clones selected for *in vitro* dry matter digestibility. Crop Sci 35:513–517

Evans RJ, Milne TA (1987) Molecular characterization of the pyrolysis of biomass. 1. Fundamentals. Energy Fuel 1:123–137

Ezeji TC, Qureshi N, Blaschek HP (2007) Bioproduction of butanol from biomass: from genes to bioreactors. Curr Opin Biotechnol 18:220–227

Fagard M, Desnos T, Desprez T, Goubet F, Refregier G, Mouille G, McCann M, Rayon C, Vernhettes S, Höfte H (2000) *PROCUSTE1* encodes a cellulose synthase required for normal cell elongation specifically in roots and dark-grown hypocotyls of Arabidopsis. Plant Cell 12:2409–2423

Faik A (2010) Xylan biosynthesis: news from the grass. Plant Physiol 153:396–402

Fincher GB (2009) Revolutionary times in our understanding of cell wall biosynthesis and remodeling in the grasses. Plant Physiol 149:27–37

Fontaine AS, Bout S, Barrière Y, Vermerris W (2003) Variation in cell wall composition among forage maize (*Zea Mays* L.) inbred lines and its impact on digestibility. Analysis of neutral detergent fiber composition by pyrolysis-gas chromatography-mass spectrometry. J Agric Food Chem 51:8080–8087

Gerhardt RL, Fritz JO, Moore KJ, Jaster EH (1994) Digestion kinetics and composition of normal and *brown midrib* sorghum morphological components. Crop Sci 34:1353–1361

Girke T, Lauricha J, Tran H, Keegstra K, Raikhel N (2004) The cell wall navigator database. A systems-based approach to organism-unrestricted mining of protein families involved in cell wall metabolism. Plant Physiol 136:3003–3008

Gorz HJ, Haskins FA, Vogel KP (1986) Inheritance of dhurrin contant in mature sorghum leaves. Crop Sci 26:65–67

Gorz HJ, Haskins FA, Morris R, Johnson BE (1987) Identification of chromosomes that condition dhurrin content in sorghum seedlings. Crop Sci 27:201–203

Grenier C, Bramel-Cox PJ, Hamon P (2001a) Core collection of sorghum. I. Stratification based on eco-geographical data. Crop Sci 41:234–240

Grenier C, Hamon P, Bramel-Cox PJ (2001b) Core collection of sorghum. II. Comparison of three random sampling strategies. Crop Sci 41:241–246

Guillaumie S, San-Clemente H, Deswarte C, Martinez Y, Lapierre C, Murgneux A, Barrière Y, Pichon M, Goffner D (2007) MAIZEWALL. Database and developmental gene expression profiling of cell wall biosynthesis and assembly in maize. Plant Physiol 143:339–363

Hamelinck C, Faaij APC (2006) Production of methanol from biomass. In: Minteer S (ed) Alcoholic fuels. Taylor and Francis, Boca Raton, FL, pp 7–50

Haney LJ, Coors JG, Lorenz AJ, Raman DR, Anex RP, Scott MP (2008) Development of a fluorescence-based method for monitoring glucose catabolism and its potential use in a biomass hydrolysis assay. Biotechnol Biofuels 1:17

Hatfield R, Ralph J, Grabber JH (2008) A potential role for sinapyl *p*-coumarate as a radical transfer mechanism in grass lignin formation. Planta 228:919–928

Holland N, Holland D, Helentjaris T, Dhugga KS, Xoconostle-Cazares B, Delmer DP (2000) A comparative analysis of the plant cellulose synthase (*CesA*) gene family. Plant Physiol 123:1313–1323

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17:754–755

Humphreys JM, Chapple C (2002) Rewriting the lignin road map. Curr Opin Plant Biol 5:224–229

Humphreys JM, Hemm MR, Chapple C (1999) New routes for lignin biosynthesis defined by biochemical characterization of recombinant ferulate 5-hydroxylase, a multifunctional cytochrome P450-dependent monooxygenase. Proc Natl Acad Sci U S A 96:10045–10050

Jung HG, Mertens DR, Buxton DR (1998) Forage quality variation among maize inbreds: in vitro fiber digestion kinetics and prediction with NIRS. Crop Sci 38:205–210

Kemsley EK (1998) Discriminant analysis and class modelling of spectroscopic data. Wiley, Chichester, UK

Kim S-J, Kim M-R, Bedgar DL, Moinuddin SGA, Cardenas CL, Davin LB, Kang C, Lewis NG (2004) Functional reclassification of the putative cinnamyl alcohol dehydrogenase multigene family in *Arabidopsis*. Proc Natl Acad Sci U S A 101:1455–1460

Kim J-S, Klein PE, Klein RR, Price HJ, Mullet JE, Stelly DM (2005) Chromosome identification and nomenclature of *Sorghum bicolor*. Genetics 169:1169–1173

Kim CM, Park SH, Il JB, Park SH, Piao HL, Eun MY, Dolan L, Han CD (2007) OsCSLD1, a cellulose synthase-like D1 gene, is required for root hair morphogenesis in rice. Plant Physiol 143:1220–1230

Knox JP (2008) Revealing the structural and functional diversity of plant cell walls. Curr Opin Plant Biol 11:308–318

Lamb JF, Haskins FA, Gorz HJ, Vogel KP (1987) Inheritance of seedling hydrocyanic acid potential and seed weight in sorghum-sudangrass crosses. Crop Sci 27:522–525

Li B-Z, Balan V, Yuan Y-J, Dale BE (2010) Process optimization to convert forage and sweet sorghum bagasse to ethanol based on ammonia fiber expansion (AFEX) pretreatment. Bioresour Technol 101:1285–1292

Liepman AH, Wilkerson CG, Keegstra K (2005) Expression of cellulose synthase-like (*Csl*) genes in insect cells reveals that *CslA* family members encode mannan synthases. Proc Natl Acad Sci U S A 102:2221–2226

Liu L, Ye XP, Womac AR, Sokhansanj S (2010) Variability of biomass chemical composition and rapid analysis using FT-NIR techniques. Carbohydr Polym 81:820–829

Lorenzana RE, Friskop Lewis M, Jung H-HG, Bernardo R (2010) Quantitative trait loci and trait correlations for maize stover cell wall composition and glucose release for cellulosic ethanol. Crop Sci 50:541–555

Lynd LR, van Zyl WH, McBride JE, Laser M (2005) Consolidated bioprocessing of cellulosic biomass: an update. Curr Opin Biotechnol 16:577–583

Mann DGJ, Labbé N, Sykes RW, Gracom K, Kline L, Swamidoss IM, Burris JN, Davis M, Stewart CN (2009) Rapid assessment of lignin content and structure in switchgrass (*Panicum virgatum* L.) grown under different environmental conditions. Bioenergy Res 2:246–256

Marita J, Vermerris W, Ralph J, Hatfield RD (2003) Variations in the cell wall composition of maize *brown midrib* mutants. J Agric Food Chem 51:1313–1321

Marsalis MA, Angadi SV, Contreras-Govea FE (2010) Dry matter yield and nutritive value of corn, forage sorghum, and BMR forage sorghum at different plant populations and nitrogen rates. Field Crops Res 116:52–57

McCarty DR, Settles AM, Suzuki M, Tan BC, Latshaw S, Porch T, Robin K, Baier J, Avigne W, Lai J (2005) Steady-state transposon mutagenesis in inbred maize. Plant J 44:52–61

Menz MA, Klein RR, Unruh N, Rooney WL, Klein PE, Mullet JE (2004) Genetic diversity of public inbreeds of sorghum determined by mapped AFLP and SSR markers. Crop sci 44:1236–1244

Miller GL (1959) Use of dinitrosalisylic acid reagent for determination of reducing sugars. Anal Chem 31:426–428

Miron J, Zuckerman E, Adin G, Solomon R, Shoshani E, Nikbachat M, Yosef E, Zenou A, Gershon Weinberg Z, Chen Y, Halachmi I, Ben-Ghedalia D (2007) Comparison of two forage sorghum varieties with corn and the effect of feeding their silages on eating behavior and lactation performance of dairy cows. Anim Feed Sci Technol 139:23–39

Mitchell RAC, Dupree P, Shewry PR (2007) A novel bioinformatics approach identifies candidate genes fo the synthesis and feruoylation of arabinoxylan. Plant Physiol 144:43–53

Mohanraj K, Gopalan A, Shanmuganathan M (2006) Genetic parameters for hydrocyanic acid content in forage sorghum. J Agric Sci 6:59–62

Mosier N, Wyman C, Dale B, Elander R, Lee YY, Holtzapple M, Ladisch M (2005) Features of promising technologies for pretreatment of lignocellulosic biomass. Biores Technol 96: 673–686

Mueller SC, Brown Jr RM (1980) Evidence for an intramembrane component associated with a cellulose microfibril synthesizing complex in higher plants. J Cell Biol 84:315–326

Murray SC, Sharma A, Rooney WL, Klein PE, Mullet JE, Mitchell SE, Kresovich S (2008) Genetic improvement of sorghum as a biofuel feedstock II: quantitative loci for stem and leaf structural carbohydrates. Crop Sci 48:2180–2193

Myton KE, Fry SC (1994) Intraprotoplasmic feruoylation of arabinoxylans in *Festuca arundinacea* cell cultures. Planta 193:326–330

Nemeth C, Freeman J, Jones HD, Sparks C, Pellny MD, Wilkinson MD, Dunwell J, Andersson AAM, Åman P, Guillon F, Saulnier L, Mitchell RAC, Shewry PR (2010) Down-regulation of the CSLF6 gene results in decreased (1,3;1,4)-β-D-glucan in endosperm of wheat. Plant Physiol 152:1209–1218

Oliver AL, Grant RJ, Pedersen JF, O'Rear J (2004) Comparison of brown midrib-6 and -18 forage sorghum with conventional sorghum and corn silage in diets of lactating dairy cows. J Dairy Sci 87:637–644

Oliver AL, Klopfenstein TJ, Grant RJ, Pedersen JF (2005a) Comparative effects of the sorghum *bmr-6* and *bmr-12* genes. I. Forage sorghum yield and quality. Crop Sci 45:2234–2239

Oliver AL, Klopfenstein TJ, Jose HD, Pedersen JF, Grant RJ (2005b) Comparative effects of the sorghum *bmr-6* and *bmr-12* genes. II. Grain yield, stover yield, and stover quality in grain sorghum. Crop Sci 45:2240–2245

Olsen AN, Ernst HA, Leggio LL, Skriver K (2005) NAC transcription factors: structurally distinct, functionally diverse. Trends Plant Sci 10:79–87

Ooka H, Satoh K, Doi K, Nagata T, Otomo Y, Murakami K et al (2003) Comprehensive analysis of NAC family genes in *Oryza sativa* and *Arabidopsis thaliana*. DNA Res 10:239–247

Palmer NA, Sattler SE, Saathoff AJ, Funnell D, Pedersen JF, Sarath G (2008) Genetic background impacts soluble and cell wall-bound aromatics in *brown midrib* mutants of sorghum. Planta 229:115–127

Palonen H, Tjerneld F, Zacchi G, Tenkanen M (2004) Adsorption of *Trichoderma reesei* CBH I and EG II and their catalytic domains on steam pretreated softwood and isolated lignin. J Biotechnol 107:65–72

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J et al (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556

Pear JR, Kawagoe Y, Schreckengost WE, Delmer DP, Stalker DM (1996) Higher plants contain homologs of the bacterial *celA* genes encoding the catalytic subunit oif cellulose synthase. Proc Natl Acad Sci U S A 93:12637–12642

Pedersen JF, Fritz JO (2000) Forages and fodder. In: Smith CW, Frederiksen RA (eds) Sorghum: origin, history, technology, and production. Wiley, New York, pp 797–810

Pedersen JF, Funell DL, Toy JJ, Oliver AL, Grant RJ (2006) Registration of seven forage sorghum genetic stocks near-isogenic for the brown midrib genes *bmr-6* and *bmr-12*. Crop Sci 46:490

Pedersen JF, Toy JJ, Funnell DL, Sattler SE, Oliver AL, Grant RA (2008) Registration of BN611, AN612, BN612 and BN613 sorghum genetic stocks with stacked *bmr-6* and *bmr-12* genes. J Plant Registr 2:258–262

Penning B, Tayengwa R, Hunter CT III, Eveland A, Vermerris W, Olek A, Koch KE, McCarty DR, Davis M, Thomas SR, McCann M, Carpita N (2009) Genetic resources for functional genomics of maize cell wall biology. Plant Physiol 153:1703–1728

Pillonel C, Mulder MM, Boon JJ, Forster B, Binder A (1991) Involvement of cinnamyl-alcohol dehydrogenase in the control of lignin formation in *Sorghum bicolor* (L.) Moench. Planta 185: 538–544

Porter KS, Axtell JD, Lechtenberg VL, Colenbrander VF (1978) Phenotype, fiber composition, and in vitro dry matter disappearance of chemically induced brown midrib (*bmr*) mutants of sorghum. Crop Sci 18:205–209

Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA, Frederick WJ Jr, Hallett JP, Leak DJ, Liotta CL, Mielenz JR, Murphy R, Templer R, Tschaplinski T (2006) The path forward for biofuels and biomaterials. Science 311:484–489

Ralph J, Grabber JG, Hatfield RD (1995) Lignin-ferulate cross-links in grasses: active incorporation of ferulate polysaccharide esters into ryegrass lignins. Carbohydr Res 275:167–178

Ralph J, Bunzel M, Marita JM, Hatfield RD, Lu F, Kim H, Schatz PF, Grabber JH, Steinhart H (2004a) Peroxidase-dependent cross-linking reactions of *p*-hydroxycinnamates in plant cell walls. Phytochem Rev 3:79–96

Ralph J, Guillaumie S, Grabber JH, Lapierre C, Barrière Y (2004b) Genetic and molecular basis of grass cell wall biosynthesis and degradability. III. Towards a forage grass idiotype. C R Biol 327:467–479

Ralph J, Hatfield RD (1991) Pyrolysis-GC-MS analysis of forage materials. J Agric Food Chem 39:1426-1437

Ralph J, Lundquist K, Brunow G, Lu F, Kim H, Schatz PF, Marita JM, Hatfield RD, Ralph SA, Christensen JH, Boerjan W (2004c) Lignins: natural polymers from oxidative coupling of 4-hydroxyphenylpropanoids. Phytochem Rev 3:29–60

Riboulet C, Fabre F, Dénoue D, Martinant JP, Lefèvre B, Barrière Y (2008) QTL mapping and candidate gene research for lignin content and cell digestibility in a top-cross of a flint maize recombinant inbreed line progeny harvested at silage stage. Maydica 53:1–9

Robinson AR, Mansfield SD (2009) Rapid analysis of poplar lignin monomer composition by a streamlined thioacidolysis procedure and near-infrared reflectance-based prediction modeling. Plant J 58:706–714

Ronquist F, Huelsenbeck JP, VanDerMark P (2005) MrBayes 3.1 Manual. http://mrbayes.csit.fsu.edu/manual.php

Rooney WL, Blumenthal J, Bean B, Mullet JE (2007) Designing sorghum as a dedicated bioenergy feedstock. Biofuels Bioprod Bioref 1:147–157

Rose JK, Braam J, Fry SC, Nishitani K (2002) The XTH family of enzymes involved in xyloglucan endotransglucosylation and endohydrolysis: current perspectives and a new unifying nomenclature. Plant Cell Physiol 43:1421–1435

Saballos A, Vermerris W, Rivera L, Ejeta G (2008) Allelic association, chemical characterization and saccharification properties of *brown midrib* mutants of sorghum (*Sorghum bicolor* (L.) Moench). Bioenergy Res 1:3–4

Saballos A, Ejeta G, Kang CH, Vermerris W (2009) A genomewide analysis of the cinnamyl alcohol dehydrogenase family in sorghum [*Sorghum bicolor* (L.) Moench] identifies *SbCAD2* as the *Brown midrib6* gene. Genetics 181:783–795

Samson D, Legeau F, Karsenty E, Reboux S, Veyrieras J-B, Just J, Barillot E (2003) GenoPlante-Info (GPI): a collection of databases and bioinformatics resources for plant genomics. Nucleic Acids Res 31:179–182

Sattler SE, Saathoff AJ, Haas EJ, Palmer NA, Funnell-Harris DL, Sarath G, Pedersen JF (2009) A nonsense mutation in a cinnamyl alcohol dehydrogenase gene is responsible for the sorghum *brown midrib6* phenotype. Plant Physiol 150:584–595

Sattler SE, Funnell-Harris DL, Pedersen JF (2010) *Brown midrib* mutations and their importance to the utilization of maize, sorghum, and pearl millet lignocellulosic tissues. Plant Sci 178:229–238

Saxena IM, Brown RM (2005) Cellulose biosynthesis: current views and evolving concepts. Ann Bot 96:9–21

Scheible W-R, Eshed R, Richmond T, Delmer D, Somerville C (2001) Modifications of cellulose synthase confer resistance to isoxaben and thiazolidinone herbicides in Arabidopsis *Ixr1* mutants. Proc Natl Acad Sci U S A 98:10079–10084

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115

Schwarz AK, Godsey CM, Luebbe MK, Erickson GE, Klopfenstein TL, Mitchell RB, Pedersen JF (2008) Forage quality and grazing performance of beef cattle grazing brown mid-rib grain sorghum residue. 2008 Nebraska Beef Reports. Lincoln, University of Nebraska

Sedlak M, Ho NWY (2004) Production of ethanol from cellulosic biomass hydrolyzates using genetically engineered *Saccharomyces* yeast capable of cofermenting glucose and xylose. Appl Biochem Biotechnol 113:403–416

Sendich E, Laser M, Kim S, Alizadeh H, Laureano-Perez L, Dale B, Lynd L (2008) Recent process improvement for the ammonia fiber expansion (AFEX) process and resulting reductions in minimum ethanol selling price. Bioresour Technol 99:8429–8434

Sewell MM, Davis MF, Tuskan GA, Wheeler NC, Elam CC, Bassoni DL, Neale DB (2002) Identification of QTLs influencing wood property traits in loblolly pine (*Pinus taeda* L.). Theor Appl Genet 104:214–222

Shen H, Yin Y, Chen F, Xu Y, Dixon R (2009) A bioinformatic analysis of NAC genes for plant cell wall development in relation to lignocellulosic bioenergy production. Bioenergy Res 2:217–232

Sibout R, Eudes A, Mouille G, Pollet B, Lapierre C, Jouanin L, Seguin A (2005) Cinnamyl alcohol dehydrogenase-C and -D are the primary genes involved in lignin biosynthesis in the floral stem of Arabidopsis. Plant Cell 17:2059–2076

Siesler HW, Ozaki Y, Kawata S, Heise HM (2002) Near-infrared spectroscopy: principles, instruments, applications. Wiley, Weinheim

Smith CW, Frederiksen RA (2000) Sorghum: origin, history, technology and production. Wiley, New York

Somerville C (2006) Cellulose synthesis in higher plants. Annu Rev Cell Dev Biol 22:53–78

Studer MH, DeMartini JD, Brethauer S, McKenzie HL, Wyman CE (2010) Engineering of a high-throughput screening system to identify cellulosic biomass, pretreatments, and enzyme formulations that enhance sugar release. Biotechnol Bioeng 105:231–238

Taylor NG, Scheible WR, Cutler S, Somerville CR, Turner SR (1999) The *irregular xylem3* locus of Arabidopsis encodes a cellulose synthase required for secondary cell wall synthesis. Plant Cell 11:769–779

Taylor NG, Laurie S, Turner SR (2000) Multiple cellulose synthase catalytic subunits are required for cellulose synthesis in *Arabidopsis*. Plant Cell 12:2529–2539

Taylor NG, Howells RM, Huttly AK, Vickers K, Turner SR (2003) Interactions among three distinct CesA proteins essential for cellulose synthesis. Proc Natl Acad Sci U S A 100:1450–1455

Tew TL, Cobill RM, Richard JEP (2008) Evaluation of sweet sorghum and sorghum × sudangrass hybrids as feedstocks for ethanol production. Bioenergy Res 1:147–152

Theander O, Åman P, Westerlund E, Andersson R, Pettersson D (1995) Total dietary fiber determined as neutral sugar residues, uronic acid residues, and Klason lignin (the Uppsala method): collaborative study. J AOAC Int 78:1030–1044

Theodorou MK, William BA, Dhanoa MS, McAllan AB, France J (1994) A simple gas production method using a pressure transducer to determine the fermentation kinetics of ruminant feeds. Anim Feed Sci Technol 48:185–197

Thompson JE, Fry SK (2001) Restructuring of wall-bound xyloglucan by transglycosylation in living plant cells. Plant J 26:23–34

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Tilley JMA, Terry RA (1963) A two-stage technique for the *in vitro* digestion of forage crops. J Br Grassl Soc 18:104–111

Van Soest PJ (1967) Development of a comprehensive system of feed analyses and its application to forages. J Anim Sci 26:119–128

Vandenbrink JP, Delgado MP, Frederick JR, Feltus FA (2010) A sorghum diversity panel biofuel feedstock screen for genotypes with high hydrolysis yield potential. Ind Crops Prod 31: 444–448

Vanholme R, Morreel K, Ralph J, Boerjan W (2008) Lignin engineering. Curr Opin Plant Biol 11:278–285

Venuto B, Kindiger B (2008) Forage and biomass feedstock production from hybrid forage sorghum and sorghum-sudangrass hybrids. Grassland Sci 54:189–196

Vermerris W (2009) Cell wall biosynthetic genes of maize and their potential for bioenergy production. In: Bennetzen J, Hake S (eds) Handbook of maize: genetics and genomics. Springer, New York, pp 741–767

Vermerris W, Nicholson R (2006) Phenolic compound biochemistry. Springer, Dordrecht

Vermerris W, Saballos A, Ejeta G, Mosier NS, Ladisch MR, Carpita NC (2007) Molecular breeding to enhance ethanol production from corn and sorghum Stover. Crop Sci 47:S145–S153

Weimer PJ, Dien BS, Springer TL, Vogel KP (2005) In vitro gas production as a surrogate measure of the fermentability of cellulosic biomass to ethanol. Appl Microbiol Biotechnol 67:52–58

Wheeler JL, Mulcahy C (1989) Consequences for animal production of cyanogenesis in sorghum forage and hay: a review. Tropical Grasslands 23:193–202

Wilson WA, Harrington SE, Woodman WL, Lee M, Sorrells ME, McCouch SR (1999) Inferences on the genome structure of progenitor maize through comparative analysis of rice, maize and the domesticated Panicoids. Genetics 153:453–473

Xin Z, Wang ML, Barkley NA, Burow G, Franks C, Pederson G, Burke J (2008) Applying genotyping (TILLING) and phenotyping analyses to elucidate gene function in a chemically induced sorghum mutant population. BMC Plant Biol 8:108–140

Xin Z, Wang ML, Burow G, Burke J (2009) An induced sorghum mutant population suitable for bioenergy research. Bioenergy Res 2:10–16

Xu Z, Zhang D, Hu J, Zhou X, Ye X, Reichel KL, Stewart NR, Syrenne RD, Yang X, Gao P et al (2009) Comparative genome analysis of lignin biosynthesis gene families across the plant kingdom. BMC Bioinformatics 10(Suppl 11):S3

Yong W, Link B, O'Malley R, Tewari J, Hunter CT, Lu CA, Li X, Bleecker AB, Koch KE, McCann MC, McCarty DR, Staiger C, Thomas SR, Vermerris W, Carpita NC (2005) Genomics of plant cell wall biogenesis. Planta 221:747–751

Zhao Q, Gallego-Giraldo L, Wang H, Zeng Y, Ding SY, Chen F, Dixon R (2010) A NAC transcription factor orchestrates multiple features of cell wall development in *Medicago truncatula*. Plant J. 63:100–114

Zugenmaier P (2001) Conformation and packing of various crystalline cellulose fibers. Prog Polym Sci 26:1341–1417

Part IV
Early Messages pertinent to Other Crops, from Saccharinae Research

Chapter 18
Comparative Genomics of Grasses: A Saccharinae-Centric View

Andrew H. Paterson, Xiyin Wang, Haibao Tang, and Changsoo Kim

Abstract Analysis of whole-genome sequences has revealed beyond doubt that all grasses are paleopolyploid, sharing at least one and probably two genome duplications since their divergence from eudicots. Comparative approaches that account for these events have begun to clarify grass–eudicot genome relationships, which were previously elusive. Within the grasses, the best-characterized clades (panicoid, oryzoid, and pooid) each include sequenced genomes that have not experienced further duplication and have evolved relatively conservatively, providing important models for many additional genomes that have experienced reduplication(s). Concerted evolution of duplicated genes is not uncommon in grass genomes and may be of remarkably long duration, particularly near the terminus of one pair of homoeologs tracing to the most recent of the pan-grass duplications. A general tendency to cyclic gain-loss of chromosome numbers, often in association with genome duplication, is evident in the Saccharinae and other cereals. Grass genomes, as well as those of eudicots are comprised of two qualitatively different "compartments," classically referred to as euchromatin and heterochromatin based on differential staining in cytological preparations, but now known to vary in many ways from one another and perhaps to provide complementary "environments" facilitating different patterns of gene and genome evolution. A small percentage of genes appear to be grass-specific, being absent from eudicot models, and are a tantalizing target for

A.H. Paterson (✉) • H. Tang • C. Kim
Plant Genome Mapping Laboratory and Departments of Crop and Soil Science, Plant Biology, and Genetics, University of Georgia, 111 Riverbend Road, Athens, GA 30602-6810, USA
e-mail: paterson@plantbio.uga.edu

X. Wang
Plant Genome Mapping Laboratory, Departments of Crop and Soil Science, Plant Biology, and Genetics, University of Georgia, 2091 South Milledge Ave, C67, Athens, GA 30602, USA

Center for Genomics and Computational Biology, School of Life Sciences, Hebei United University, Tangshan, Hebei 063009, China
e-mail: wang.xiyin@gmail.com

A.H. Paterson (ed.), *Genomics of the Saccharinae*, Plant Genetics and Genomics: Crops and Models 11, DOI 10.1007/978-1-4419-5947-8_18,

early functional analysis but caution is especially warranted in selecting such genes for study. Despite rapid progress in recent years, much of Saccharinae, and indeed Poaceae, diversity remains inadequately explored. A host of additional questions about the evolution of this clade are likely to become answerable with accelerating progress in genome analysis, both adding to fundamental knowledge of botanical diversity and further enhancing the contributions of members of the clade to production of food, feed, fiber, and fuel.

Keywords Paleopolyploidy • Concerted evolution • Euchromatin • Heterochromatin • Chromosome number evolution • Lineage-specific genes • Botanical diversity

1 The Common Ancestor of Grasses

The major Poaceae (grass) lineages are thought to have diverged from a common ancestor about 40–50 MYA. The large number of Poaceae crops and their seemingly conservative evolution has made them a primary beneficiary of comparative genomics studies, utilizing information and tools from relatively small and well-studied genomes such as those of rice and sorghum in the study and improvement of larger and more complex genomes such as wheat and sugarcane.

Comparison of the genome sequences of sorghum, rice, and Brachypodium, each unaffected by additional genome duplications since their divergence from a common ancestor, have shed much light on the evolution of the grass lineage. We now know that there have been at least two large-scale, perhaps whole-genome, duplications in the common ancestor of most (probably all) grasses, that much of process of adaptation to the duplicated state was shared among these and perhaps all grass lineages, and that in a subset of lineages additional more recent duplications have profoundly affected genome content and gene linkage relationships. Nonetheless, key features of genome organization have been preserved in parallel in these lineages since their divergence about 40–50 MYA, including broad patterns of genome organization and singular evolutionary mechanisms and rates that distinguish particular genomic regions. Here we explore the evolutionary history of the grass lineage.

2 The Pan-Cereal Genome Duplication About 70 MYA

The discovery of multiple ancient polyploidization events in *Arabidopsis* (Vision et al. 2000; Simillion et al. 2002; Bowers et al. 2003b) foreshadowed the finding that grasses, too, had experienced ancient genome duplication. Hints of such duplication were known even before the sequencing of the first grass genomes—it had long been known that rice chromosomes occasionally paired with seemingly incorrect partners (Lawrence 1931), and had been shown by RFLP mapping that rice

chromosomes 1–5 (Kishimoto et al. 1994) and 11–12 (Nagamura et al. 1995) as well as several pairs of sorghum linkage groups (Chittenden et al. 1994) each contained duplicated gene pairs in what appeared to be collinear orders.

Initial analysis of an early rice genomic shotgun sequence suggested a widespread propensity for gene duplication that was consistent with a large-scale event estimated to have been perhaps 40–50 MYA (Goff et al. 2002). Nonrandom patterns of gene duplication based on identification by blastp (Altschul et al. 1990) of high-quality nonself matches to rice genes in their genomic orientation along "pseudomolecules" (putative chromosomes) clarified the structure of this ancient duplication event, revealing nine nonoverlapping "duplicated blocks" to account for 61.9 of the rice transcriptome (excluding repetitive genes), with individual blocks ranging from 1.8 to 13.8 of the transcriptome.

Several lines of evidence showed that this duplication event preceded the divergence of the major grass lineages, i.e., occurred in a grass common ancestor. Prior to the availability of genomic sequences for other grasses, analysis of gene trees including one pair of syntenic duplicated rice genes, the best-matching sequence from a variety of other grasses, and *Physcomitrella* as an outgroup, suggested that rice duplication is more recent than its divergence from its closest nongrass relative studied (Musa), but more ancient than its divergence from panicoids (*Sorghum*) and pooids (*Hordeum*) (Paterson et al. 2004b), later extended to also include chloridoids (Kim et al. 2009). While substantial variation in evolutionary rates among taxa (Tang et al. 2008a) cast some doubt on the robustness of gene tree inferences, "one-to-one" genome alignments among sorghum, rice and Brachypodium in all combinations also supported the conclusion that genome duplication in a common ancestor predated the divergence of the major grass lineages (Tang et al. 2008a, b, 2010a). A controversy about whether this genome duplication reflected "ancient aneuploidy," i.e., affected only part of the genome (Vandepoele et al. 2003), was eventually attributed to unduly stringent statistical thresholds (Paterson et al. 2005). Post-duplication gene loss was generally orthologous in rice and sorghum (Paterson et al. 2009b), suggesting that adaptation of the lineage to the polyploid state was largely complete prior to radiation of the grass lineages, and that patterns of gene retention/loss in these grass genomes are likely to be predictive of others. Fossil evidence suggesting that the grass radiation had already occurred by 70 MYA (Prasad et al. 2005) raises the possibility that grass molecular clocks (even based on large genome-wide gene samples) may be generally underestimating the antiquity of these events.

2.1 An Earlier Pan-Grass Genome Duplication May Have Contributed Further to Monocot–Dicot Divergence

Several studies subsequent to discovery of the "rho" duplication have hinted that additional monocot duplications may have predated rho (Zhang et al. 2005; Jaillon et al. 2007; Salse et al. 2008). The combination of a visually intuitive approach with a gene-based multi-collinearity searching algorithm MCscan (Tang et al. 2008b), provided evidence of additional duplication blocks of deeper hierarchy, covering at

least 20 of the cereal transcriptome. Specifically, merger of the gene composition and order from corresponding rho-duplicated blocks in both sorghum and rice permitted the inference of an approximate pre-rho genome composition prior to the rho duplication (Tang et al. 2010a), intended to computationally reverse post-rho gene loss and thereby increase the sensitivity of subsequent analysis. This order can be only an approximation, since the ancestral positions of the intervening singleton genes between consecutive pairs of paralogs cannot be precisely determined. The pre-rho reconstructed gene orders were compared among themselves, revealing collinear patterns of correspondence that involve all nine major rho-blocks, some one-to-one while others involve more than two rho-blocks. To facilitate further analysis, we curated a second list of eight large "sigma-blocks" that have retained collinearity following rho, containing a total of 4,168 nodes covering 5,747 rice genes and 5,738 sorghum genes (~20 of the rice and sorghum transcriptomes). It is difficult to enumerate all patterns of sigma collinearity, since some duplicated regions become highly degenerate during post-WGD diploidization, creating gene orders that are largely reciprocal or sometimes complementary (Van de Peer 2004; Freeling et al. 2008). Relationships between some degenerate segments can still be identified through transitive comparisons of grapevine and rice genomes, but there is little remaining intra-genomic sigma correspondence between rice segments.

Based on a molecular clock of 6.5×10^{-9} synonymous substitutions per synonymous site per year (Gaut et al. 1996), the sigma duplication(s) are estimated to have occurred approximately 130 MYA (Tang et al. 2010a).

2.2 *Monocot–Dicot Parallels*

Similarities between monocot and eudicot genomes resulting from common ancestry have been obscured by many rounds of paleopolyploidy and numerous genome rearrangements (Liu et al. 2001; Jaillon et al. 2007). K_s-based "distances" between sigma duplicates (median 1.72) are less than those between grape–cereal orthologs (median K_s 1.95: $P = 4.8 \times 10^{-24}$, student's *t*-test) indicating that the sigma event(s) largely if not entirely preceded the monocot–dicot divergence. However, differences in lineage specific mutation rates confound interpretation of K_s values (Tang et al. 2008a) and we reemphasize that our divergence time estimates must be considered rough approximations.

Recurring polyploidy events pose significant challenges when comparing monocot and eudicot genomes because of the degeneration caused by independent gene fractionation (or "diploidization") following several rounds of paleopolyploidy in each lineage. A hierarchical clustering approach partially circumvents such difficulties to identify synteny across grape and rice (Tang et al. 2010b). Briefly, the chromosomes were first cut into small segments and comparisons were made between every pair of rice and grape segments. Duplicated segments retained in grape following the eudicot "gamma" hexaploidy event (Jaillon et al. 2007), and homologous segments retained in rice following at least two rounds of duplication (rho and sigma), contain 38 "putative ancestral regions" (PAR) that interleave multiple grape

and rice genomic regions collectively covering ~70 of each genome. By consolidating much of the redundancy in each genome, the PARs create syntenic blocks with less ambiguity and in most cases show association between one gamma block and one sigma block. The individual PARs derived from grapevine-rice comparisons offer an independent and important validation of the sigma blocks identified through cereal intra-genomic comparisons, also revealing some "ghost duplications" (Vandepoele et al. 2002) that we failed to identify through intra-genomic comparisons (due to reciprocal gene losses in largely complementary fashion).

When a particular PAR is scrutinized, syntenic relationships among the clustered regions are more informative than analyzing any individual pair of syntenic segments that contribute to the PAR. In 22 of the 38 PARs, grape–rice collinearity is clear, which allows us to evaluate the level of redundancies (reflecting the number of genome duplication events) observable in both lineages. Among the 22 PARs, twelve- are threefold redundant in grapevine, consistent with hexaploidy (Jaillon et al. 2007). The level of redundancy in rice is less clearcut, ranging from as little as twofold (one PAR) to sevenfold (three PARs) and eightfold (five PARs), suggesting more than two, perhaps three, rounds of WGD. In other words, "sigma" may be a composite of two ancient events that we presently lack the resolution to distinguish from one another.

3 Lineage-Specific Grass Genome Duplications and Their Consequences

After the two pan-grass duplications and radiation of the major grass lineages, several have experienced additional genome duplications with profound impact on genome organization and gene function.

For example, *Sorghum* and *Zea* (maize, the leading US crop with a farm-gate value of $15–20 billion/year, in the Andropogonae but just outside the Saccharinae) diverged from a common ancestor ~12 MYA (Gaut et al. 1997; Swigonova et al. 2004a) versus ~42 MYA for rice and the maize/sorghum lineage (Paterson et al. 2004b). Despite 30 million years less divergence, the genome structure of sorghum much more closely resembles that of rice than maize (Bowers et al. 2003a). *Zea* has experienced one whole-genome duplication since its divergence from *Sorghum* (Swigonova et al. 2004b), with associated rearrangement and gene loss resulting in substantial restructuring of its genome relative to sorghum, rice, and other grasses.

Saccharum (sugarcane), arguably the most important biofuels crop worldwide, valued at ~$30 billion including $1 billion/year in the USA, is even more closely related to sorghum than maize, perhaps sharing common ancestry as little as 5–9 MYA (Sobral et al. 1994; Jannoo et al. 2007), retaining similar gene order (Ming et al. 1998), and even producing viable progeny in some intergeneric crosses (Dewet et al. 1976). Yet, in less time than *Zea* took to experience one whole-genome duplication, *Saccharum* experienced at least one independent whole-genome duplication (Ming et al. 1998).

Miscanthus, a promising bioenergy crop for temperate latitudes (Heaton et al. 2008), is thought to have diverged from *Saccharum* still more recently than their common ancestor did from sorghum. *Miscanthus* species have a basal set of 19 chromosomes ($2n = 38$, and 38 or 76 for spp. *sinensis* and *sacchariflorus*, respectively), versus the ten that is characteristic of many Saccharinae including *Saccharum*. One attractive hypothesis to explain the transition from 10 to 19 chromosomes is that *Miscanthus*, like *Saccharum*, may have experienced a polyploidization since its divergence from sorghum. However, its basal chromosome number of 19 suggests that unlike those of *Saccharum* which is largely autopolyploid, *Miscanthus* homologs may have diverged sufficiently that they no longer normally pair with one another (perhaps including a chromosomal fusion to get from 20 to 19). In view of their close relationship, perhaps *Miscanthus* and *Saccharum* even shared a genome doubling, diverging later. Other options of polyploidy followed by chromosome number reduction as is evident in several *Sorghum* species and suggested in maize (Tang et al. 2008a), are contra-indicated by recent genetic mapping studies (Kim et al. 2012). Genetic mapping of *Miscanthus* promises to clarify the relationship of its chromosomes to those of sorghum and *Saccharum*. Knowledge of the mechanisms, levels, and patterns of evolution of genome size and structure in this curious group will help to reveal the path by which the sorghum genome has arrived at its present state, also laying the foundation for further study of sugarcane and other economically important members of the group.

One can point to many additional examples of naturally occurring genome duplication in grass lineages, as well as some human-mediated examples that have led to significant crops such as triticale. Relatively recent duplications such as those experienced in wheat, may provide a valuable complement to the more ancient maize, sugarcane, rho and sigma duplications in unraveling correlated gene fates and formulating hypotheses about adaptation of genomes to the duplicated state.

4 Evolution of Paleoduplicated Genes and Genomes

The duplication of a genome and its constituent genes opens doors to otherwise-infrequent evolutionary mechanisms and events. First and foremost, both genetic theory and repeated observation suggest that following genome duplication, the vast majority of duplicated gene will be lost. In the case of grasses, the massive gene loss following genome duplication(s) was largely complete prior to the divergence of the major lineages from a common ancestor, as rice, sorghum, and Brachypodium, genomes that have not been reduplicated, have closely correlated patterns of gene retention/loss. Even in these relatively well-conserved genomes, there have been occasional divergent losses of homoeologs since lineage divergence, and which are not randomly distributed across gene functional groups, suggesting the action of selection (Paterson et al. 2009a). In genomes such as maize, sugarcane, wheat, and others that have experienced additional duplication or triplication, new

lineage-specific "waves" of gene loss can dramatically alter ancestral gene linkage relationships and offer an opportunity for selection to act on the gene repertoire.

The presence in a genome of duplicated copies of most/all genes, that may be virtually identical to one another shortly after the duplication event, raises a fascinating question—when do the newly formed duplicate genes establish independent evolution? For most such genes, independence is achieved when the alternative copy is mutated and lost or degraded. However, what of the substantial population of genes for which both duplicated copies are retained?

The mere presence of two copies of a DNA sequence in the same genome raises the possibility that gene conversion, often accompanied by crossing-over, can homogenize genetic variation to render similar DNA sequences identical (Galtier 2003). One model for recombination suggests that gene conversion may be explained by repair of unmatched bases during the formation of heteroduplex DNA (Holliday 1966). Gene conversion is often involved in homogenization of small tracts of paralogous DNA sequences, usually between several and several hundred base pairs (Petes and Symington 1991), whereas the homogenization of larger tracts of DNA is generally believed to involve crossing-over (Szostak and Wu 1980). Traditionally, gene conversion was used to describe the evolution of rRNA (Brown et al. 1972) and histone genes (Ohta 1984), both occurring in tandem clusters having tens of copies in an organism. Gene conversion has also been proposed to affect the evolution of various multigene families (Sawyer 1989; White and Crowther 2000; Mondragon-Palomino and Gaut 2005).

Concerted evolution of duplicated genes is not uncommon in grass genomes, and may be of remarkably long duration. Concerted evolution of duplicated genes may alter the topology of "gene trees" in predictable ways, for example rendering ancient paralogs within the same nucleus more similar to one another than orthologs in closely related taxa. Rice was the first plant species for which two divergent subspecies were largely sequenced, revealing appreciable gene conversion between 70 million-year-old duplicates in the past 400,000 years since their divergence, with a gradual progression toward independent evolution of older paralogs (Wang et al. 2007). Domain-encoding sequences are more frequently converted than nondomain sequences, suggesting a sort of circularity—that sequences conserved by selection may be further conserved by relatively frequent conversion.

With the release of the sorghum genome sequence, it became possible to look more deeply, investigating concerted evolution of genes in both sorghum (using rice as outgroup) and rice (using sorghum as outgroup) in the ~50 million years since their divergence (Wang et al. 2009). An estimated 18 and 13 of rice and sorghum duplicated genes, respectively, have been affected by concerted evolution after rice–sorghum divergence. Gene concerted evolution occurs most frequently in gene-rich euchromatic regions, perhaps suggesting that sequences and their divergence contribute to restriction of ectopic recombination. Though converted paralogs are more similar to one another than nonconverted ones, elevated nucleotide differences between rice–sorghum orthologs indicates that converted genes have evolved at a faster rate than other genes, implying that concerted evolution acts as an accelerating rather than a conservative element.

Concerted evolution has not been evenly distributed across the grass genomes. Comparison of rice, sorghum, maize, and *Brachypodium* genomes revealed that one paleoduplicated chromosome pair has experienced illegitimate recombination for tens of millions of years, which has been temporally restricted in a stepwise manner, producing chromosome structural stratification. These strata formed independently in different grass lineages, with their similarities (low K_s of paleoduplicated genes) preserved in parallel for millions of years since divergence of these lineages. The pericentromeric region of this homoeologous chromosome pair accounts for 2/3 of gene content differences between the modern chromosomes. Both intriguing and perplexing is a distal chromosomal region with the greatest DNA similarity between surviving duplicated genes but the highest concentration of lineage-specific gene pairs found anywhere in these genomes and a significantly elevated gene evolutionary rate. Intragenomic similarity near this chromosomal terminus may be important in hom(e)ologous chromosome pairing. Chromosome structural stratification, together with enrichment of autoimmune response-related (NBS-LRR) genes and accelerated DNA rearrangement and gene loss, confer striking resemblance of this grass chromosome pair to the sex chromosomes of other taxa.

5 Chromosome Number Evolution

Chromosome number, together with chromosome recombinational length, determines the range of new allele combinations that can arise in successive generations. What happens to chromosome number during adaptation of a lineage to genome duplication? In extreme cases such as Arabidopsis, in which the modern $n = 5$ karyotype has been shaped by two whole-genome duplications and one whole-genome triplication, it is clear that there must have been a general tendency toward chromosome number reduction following duplication (otherwise, the pretriplication ancestor would have had only 0.4 chromosomes).

A general tendency to cyclic gain-loss of chromosome numbers, often in association with genome duplication, is evident in the Saccharinae and other cereals. A particularly striking case is in *Zea mays* (maize), which experienced a whole-genome duplication in the ~10–15 million years since its divergence from a common ancestor shared with sorghum (Swigonova et al. 2004b), *but today has the same chromosome number as sorghum* (10). Comparative mapping shows most individual sorghum chromosomes to correspond to two different chromosome arms in maize, suggesting that about ten chromosome fusions have followed the maize genome duplication.

However, the notion of cyclic gain-loss of chromosome in association with genome duplication also fails to explain numerous karyotypes, at least so far. We introduced earlier in this chapter the case of *Saccharum* and *Miscanthus*, thought to have diverged from a common ancestor more recently than their common ancestor did from sorghum. Limaize, both have experienced genome duplication since divergence from sorghum. Genetic mapping shows clearly (Ming et al. 1998) that the

10-chromosome homologous series that is characteristic of *Saccharum robustum* and *S. officinarum*, are largely syntenic with entire sorghum chromosomes, with only partial differentiation into "homoeologs," i.e., with duplicated chromosomes still pairing and recombining in most if not all possible combinations. Its albeit still-unfolding genetic map (Kim and Jakob 2012) generally indicate that two *Miscanthus* chromosomes correspond to each sorghum chromosome, again suggesting little condensation although we anticipate that one such condensation will eventually be found (explaining its basal $n=19$ rather than 20).

Further exceptions to the cyclic gain-loss hypothesis appear to exist in the parasorghums, in which several taxa have $n=5$ karyotypes. The $n=5$ species of *Sorghum* are clearly derived in the phylogeny (Spangler et al. 1999), pointing to a secondary reduction in chromosome number (also see Kellogg Chap. 1 this volume). Moreover, only a recent origin of the $n=5$ karyotype would be consistent with the finding (Paterson et al. 2004b, 2009a) that the *Sorghum bicolor* ($n=10$) genome has not experienced genome duplication or paleopolyploidy in 70 million years or more, ruling out the possibility of formation of its $n=10$ karyotype from $n=5$ sorghums. Two studies (Garber 1950; Price et al. 2005) found that chromosomes of the $x=5$ species were considerably larger than those of the $x=10$ taxa; the $x=5$ sorghums also have a higher 2C DNA content than the $x=10$ species (Price et al. 2005). These observations would all be explained if $x=5$ represents chromosomal condensations—however unlike the maize case, these condensations would come 70 million years after the most recent polyploidy.

6 Tales of Two Genomes

Careful comparison of the first two grass genomes sequenced, rice and sorghum, separated by about 50 million years, illustrated a principle that is holding true for the genomes of many additional angiosperms.

Remarkably, about 62 of the sorghum genome is almost recombinationally inert. A genetic map based on a cross of BTx623 (the sequenced genotype) and its wild relative *S. propinquum* (also physically mapped; Bowers et al. 2005) including 2,512 loci that defined 61.5 of the recombination events in the underlying population (Bowers et al. 2003a), could be aligned with the genome sequence based on sequences for 2,050 probes that identified about 90 of the loci (some probes hybridizing to multiple loci). The genome could be clearly partitioned into a recombination-rich fraction estimated to comprise 252 mbp of DNA and accounting for 1,025.2 cM (97) of the length of the genetic map; and a recombination-poor fraction comprising about 460 mbp of DNA and accounting for 34 cM (3) of the length of the genetic map.

Cytologically identified heterochromatin (Chen et al. 2002; Jiao et al. 2005) corresponds very closely to the recombinationally inert region(s) of the sorghum genome. The locations of heterochromatin in the sorghum physical map

(Bowers et al. 2005) and sequence (Paterson et al. 2009b) were estimated based on relative distance from the centromere, assigning approximate base pair locations without accounting for sequence gaps.

The euchromatin–heterochromatin distinction is associated closely with not only differences in recombination rate, but also with differential abundance of low copy-number genes and gene-associated DNA transposons (frequent in euchromatin), and high-copy number genes and relatively "old" LTR retroelements (frequent in heterochromatin).

LTR retrotransposons comprised 55 of the sorghum genome, and about 98 of the total repetitive DNA in the genome, accounting for virtually all of the difference in physical size of the sorghum and rice genomes.

LTR retroelements of different ages showed striking differences in genomic distribution (Paterson et al. 2009b). Relatively young (i.e., less than 0.1 MYA) LTR insertions are approximately uniformly distributed across the genome, while older insertions are heavily concentrated in heterochromatic regions immediately surrounding the centromeres. This closely parallels a proposal based on comparison of the sorghum physical map to the rice sequence, that selection against gene and genome structural mutation is much stronger in the gene-rich euchromatin than in the repeat-rich heterochromatin (Bowers et al. 2005).

The genomic distribution of regions in which paleoduplication could/could not be discerned added further weight to the notion that the sorghum genome is comprised of two highly divergent "components" with remarkably different properties and evolutionary histories. Discernible paleoduplication was almost completely confined to the "euchromatin," consistent with the hypothesis (Bowers et al. 2005) that genomic rearrangements are usually deleterious, thus more likely to persist in nonrecombinogenic regions by virtue of "Muller's ratchet" (Muller 1964). Moreover, selection may actually favor the rapid restructuring of heterochromatic regions that contain centromeres (Bowers et al. 2005). A high concentration of rice genes duplicated by ancient polyploidy fall near K_s 0.85 (Paterson et al. 2004a). Rice gene pairs with K_s 0.2–0.6 tend to be concentrated in peri/centromeric regions, suggesting that a substantial restructuring of these regions began shortly after polyploidization and lasted until about 16 MYA.

In partial summary, the sorghum genome and those of other grasses, as well as dicot plants (Schmutz et al. 2010) are comprised of two qualitatively different "compartments." The "euchromatin," classically defined based on differential staining in cytological preparations, tends to be distal to the chromosome, is rich in low-copy genes and indeed contains most of the genes in a genome, accounts for most of the recombination, and is resistant to structural change both on a small scale (retroelement insertions) and a large scale (chromosomal rearrangements). The "heterochromatin" tends to be central to the chromosome, is relatively gene-poor although enriched in gene families of more than a few members, may comprise much of the genomic DNA but is largely repetitive elements and recalcitrant to recombination, and experiences relatively rapid structural evolution and DNA sequence turnover (SanMiguel et al. 1996). The physical locations of these respective domains correspond closely in chromosomes that trace to the ~70 million year old genome

duplication affecting most if not all grasses, suggesting that the "compartmentation" is of extremely long duration.

One could imagine that each compartment may be essential to the evolution of particular gene families or gene arrangements—for example, the propagation of high-copy gene families being facilitated by the heterochromatin. Purifying selection acting on low-copy genes may be facilitated by the high levels of recombination of the euchromatin. The heterochromatin, in turn, is more conducive to the formation of "coadapted gene complexes" (Lande 1975) which have come to include *cis*-linkage of alleles at different loci that collectively confer an adaptive phenotype. For example, QTLs for many additional traits (beyond flowering) are closely associated with the major determinant of short-day flowering in sorghum, *Ma1*, including tillering, rhizomatousness and "ratooning" or regrowth (Paterson et al. 1995b); seed size (Paterson et al. 1995a), and leaf length and width (Paterson AH, unpublished). Studies by the author's lab delineated *Ma1* to a "bin" of about 1 cM, which also includes *dw2*, the sorghum gene with the largest phenotypic effect on plant height (Lin 1998). The sorghum genome sequence reveals this particular 1 cM to be a remarkable 36 million base-pairs (mbp) in length, with 50-fold less recombination than the genome-wide average of 0.7 mbp/cM. Increased height affords a competitive advantage in light interception and seed dispersal, and short-day flowering synchronizes seed development with water availability in the semi-arid tropical centers of origin of most Poaceae (Harper 1977). Favorable alleles at different genes that conferred both optimal height and flowering time to the same progeny by virtue of suppressed recombination might have become fixed more quickly than independently segregating alleles conferring these attributes.

7 Comparative Gene Inventories

Most angiosperms share most genes, and the Saccharinae are no exception. The number and sizes of sorghum gene families are similar to those for *Arabidopsis*, rice and poplar (sorghum sequence), with a total of 9,503 (58 of) families shared among all four species and 15,225 (93) of sorghum gene families overlapped with at least one of the three other species. Nearly 94 of the high confidence sorghum genes (25,875/27,640) have orthologs in rice, *Arabidopsis*, and/or poplar, and together these gene complements define 11,502 ancestral angiosperm gene families represented in at least one contemporary grass and rosid genome.

However, 3,983 (24) of the gene families had members only from sorghum and rice; and 1,153 (7) appear to be unique to sorghum, being absent from rice or diverged beyond recognition. One might expect lineage-specific genes to be relatively frequent in angiosperms due to recurring polyploidization/diploidization cycles (Tang et al. 2008a) that do increase regulatory and may increase morphological complexity (Freeling and Thomas 2006); and to transposition of gene segments that may evolve new genes (Jiang et al. 2004; Bennetzen 2005), albeit rarely (Juretic et al. 2005).

Lineage-specific genes are a tantalizing target for early functional analysis because they may perhaps relate to features that distinguish closely related taxa, but caution is especially warranted in selecting genes for analysis. Rapid gene evolution may be due to a lack of structural or functional constraint or to strong positive selection for functional divergence. Genes under strong positive selection in *Drosophila*, mammals, and several other species are vital to reproductive success, cell–cell recognition, and cellular response to pathogens (e.g., Yang et al. 2000; Swanson et al. 2001a, b). However, estimating frequencies of lineage-specific genes is inherently error-prone, the variance of a difference (lineage-specific genes) being the sum of variances of the respective parameters (gene numbers in each species). Transposable-element associated genes have been a major contributor to inflated gene number estimates and false positive inferences of lineage-specific angiosperm genes (Bennetzen et al. 2004). Whole-genome shotgun sequencing approaches (often including many gaps), and occasional errors in accurate inferences of intron–exon boundaries, each may artifactually truncate or elongate gene models. Indeed, gene annotations of virtually all angiosperm genomes, even *Arabidopsis* which has been scrutinized by hundreds of scientists for nearly a decade, remain works in progress and different annotations of even one genome invariably differ. While much may be learned from functional analyses of lineage-specific genes, their careful manual annotation should precede investments in their analysis. Each additional genome available to study using comparative approaches further improves power to expose false-positive cases of apparent lineage-specific genes, for example by gross phylogenetic incongruities in inferred gene sets (Paterson et al. 2009a).

8 Looking Forward

Many members of the Saccharinae group of cereals (see Kellogg, Chap. 1, this volume) have large and complex genomes that derive great benefit from a closely related genomic model. This interesting group shows sixfold variation in genome size among closely related species with the same chromosome number (for example, *S. bicolor* and *S. propinquum* versus *S. nitidum*) (Price et al. 2005); an apparent reduction in chromosome number from the ancestral $2n = 20$ to $2n = 10$ in most parasorghums (Spangler et al. 1999); at least two chromosome doublings in *Saccharum* since its divergence from the remainder of the group (Ming et al. 1998); and both natural (*Sorghum halepense*: Paterson et al. 1995b) and human-mediated polyploidization (*Saccharum* cultivars: Ming et al. 1998).

The small genome of sorghum has long been an attractive model for advancing understanding of the structure, function, and evolution of grass genomes. Sorghum is representative of many grasses of tropical origin in that it has “C4” photosynthesis, using complex biochemical and morphological specializations to improve carbon assimilation at high temperatures. By contrast, rice, the first grass genome to be sequenced, is more representative of temperate grasses, using “C3” photosynthesis. Sorghum and rice each share the distinction, as does Brachypodium (Initiative 2010)

of a lack of genome duplication since an event in a common ancestor of the three about 70 MYA (Paterson et al. 2004b). However, sorghum is much more closely related than rice to Saccharinae crops such as sugarcane, and other Andropogonae crops such as maize. Indeed, *Saccharum* (sugarcane), arguably the most important biofuels crop worldwide, valued at ~$30 billion including about $1 billion/year in the USA, may have shared ancestry with sorghum as little as 5–9 MYA (Sobral et al. 1994; Jannoo et al. 2007), retains similar gene order (Ming et al. 1998), and even produces viable progeny in some intergeneric crosses (Dewet et al. 1976). *Zea* has undergone one whole-genome duplication since its divergence from *Sorghum* (Swigonova et al. 2004b), and *Saccharum* has undergone at least two (Ming et al. 1998). These merits of sorghum have stimulated invigorated activity in its functional genomics, described elsewhere in this volume, that is expected to advance progress toward its role as a Saccharinae model.

Despite rapid progress in recent years, much of Saccharinae, and indeed Poaceae, diversity remains inadequately explored. What is the true relationship of the ten sorghum chromosomes to the five of its parasorghum relatives? Or, to the 19 of *Miscanthus*? Or even the ten of Saccharum, which after nearly 20 years of effort still lack any one genetic map that is truly complete!? What have been the structural and functional consequences of the joining of *S. bicolor* and *S. propinquum* in a common nucleus, the first genome duplication in this lineage in 70 million years, to form a new species, *S. halepense*? What is the nature of the genomes of major grass clades that are not yet adequately explored, for example the chloridoids?

While several fully sequenced genomes (rice, maize, Setaria, Brachypodium) suffice for comparative analyses within the Saccharinae, broader questions across the Poaceae are increasingly in need of an outgroup for the grasses generally. For example, what is the true relationship of the pan-grass genome duplication (Paterson) to the origins of Poaceae diversity? Several other Poales families are economically important (Bromeliaceae, Cyperaceae) but to date have only limited DNA sequence information. Several more distant monocots (palm, banana) are being sequenced but their value as models is constrained by their evolutionary distance. The identification and sequencing of a Poaceae outgroup may offer singular opportunities to grass comparative biology.

References

Altschul S, Gish W, Miller W et al (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Bennetzen JL (2005) Transposable elements, gene creation and genome rearrangement in flowering plants. Curr Opin Genet Dev 15:621–627

Bennetzen JL, Coleman C, Liu RY, Ma JX, Ramakrishna W (2004) Consistent over-estimation of gene number in complex plant genomes. Curr Opin Plant Biol 7:732–736

Bowers JE, Abbey C, Anderson S, Chang C, Draye X, Hoppe AH, Jessup R, Lemke C, Lennington J, Li Z, Lin YR, Liu SC, Luo L, Marler BS, Ming R, Mitchell SE, Kresovich S, Schertz KF, Paterson AH (2003a) A high-density genetic recombination map of sequence-tagged sites for

sorghum, as a framework for comparative structural and evolutionary genomics of tropical grains and grasses. Genetics 165:367–386
Bowers JE, Chapman BA, Rong JK, Paterson AH (2003b) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 422:433–438
Bowers JE, Arias MA, Asher R, Avise JA, Ball RT, Brewer GA, Buss RW, Chen AH, Edwards TM, Estill JC, Exum HE, Goff VH, Herrick KL, Steele CLJ, Karunakaran S, Lafayette GK, Lemke C, Marler BS, Masters SL, McMillan JM, Nelson LK, Newsome GA, Nwakanma CC, Odeh RN, Phelps CA, Rarick EA, Rogers CJ, Ryan SP, Slaughter KA, Soderlund CA, Tang HB, Wing RA, Paterson AH (2005) Comparative physical mapping links conservation of microsynteny to chromosome structure and recombination in grasses. Proc Natl Acad Sci USA 102:13206–13211
Brown DD, Wensink PC, Jordan E (1972) A comparison of the ribosomal DNA's of *Xenopus laevis* and *Xenopus mulleri*: the evolution of tandem genes. J Mol Biol 63:57–73
Chen MS, Presting G, Barbazuk WB, Goicoechea JL, Blackmon B, Fang FC, Kim H, Frisch D, Yu YS, Sun SH, Higingbottom S, Phimphilai J, Phimphilai D, Thurmond S, Gaudette B, Li P, Liu JD, Hatfield J, Main D, Farrar K, Henderson C, Barnett L, Costa R, Williams B, Walser S, Atkins M, Hall C, Budiman MA, Tomkins JP, Luo MZ, Bancroft I, Salse J, Regad F, Mohapatra T, Singh NK, Tyagi AK, Soderlund C, Dean RA, Wing RA (2002) An integrated physical and genetic map of the rice genome. Plant Cell 14:537–545
Chittenden LM, Schertz KF, Lin YR, Wing RA, Paterson AH (1994) A detailed Rflp map of *Sorghum bicolor* X *S. propinquum*, suitable for high-density mapping, suggests ancestral duplication of sorghum chromosomes or chromosomal segments. Theor Appl Genet 87:925–933
Dewet JMJ, Gupta SC, Harlan JR, Grassl CO (1976) Cytogenetics of introgression from Saccharum into *Sorghum*. Crop Sci 16:568–572
Freeling M, Thomas BC (2006) Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. Genome Res 16:805–814
Freeling M, Lyons E, Pedersen B, Alam M, Ming R, Lisch D (2008) Many or most genes in Arabidopsis transposed after the origin of the order Brassicales. Genome Res 18:1924–1937
Galtier N (2003) Gene conversion drives GC content evolution in mammalian histones. Trends Genet 19:65–68
Garber ED (1950) Cytotaxonomic studies in the genus *Sorghum*. Univ Calif Publ Bot 23:283–362
Gaut BS, Morton BR, McCaig BC, Clegg MT (1996) Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene Adh parallel rate differences at the plastid gene rbcL. Proc Natl Acad Sci USA 93:10274–10279
Gaut BS, Clark LG, Wendel JF, Muse SV (1997) Comparisons of the molecular evolutionary process at rbcL and ndhF in the grass family (Poaceae). Mol Biol Evol 14:769–777
Goff SA, Ricke D, Lan TH, Presting G, Wang RL, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchinson D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong JP, Miguel T, Paszkowski U, Zhang SP, Colbert M, Sun WL, Chen LL, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu YS, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T, Oliphant A, Briggs S (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp japonica). Science 296:92–100
Harper J (1977) Plant population biology. Academic Press, London
Heaton EA, Dohleman FG, Long SP (2008) Meeting US biofuel goals with less land: the potential of *Miscanthus*. Glob Chang Biol 14:2000–2014
Holliday R (1966) Studies on mitotic gene conversion in Ustilago. Genet Res 8:323–337
Initiative TIB (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature 463:763–768
Jaillon O, Aury JM et al (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463–467

Jannoo N, Grivet L, Chantret N, Garsmeur O, Glaszmann JC, Arruda P, D'Hont A (2007) Orthologous comparison in a gene-rich region among grasses reveals stability in the sugarcane polyploid genome. Plant J 50:574–585

Jiang N, Bao ZR, Zhang XY, Eddy SR, Wessler SR (2004) Pack-MULE transposable elements mediate gene evolution in plants. Nature 431:569–573

Jiao Y, Jia P, Wang X, Su N, Yu S, Zhang D, Ma L, Feng Q, Jin Z, Li L, Xue Y, Cheng Z, Zhao H, Han B, Deng XW (2005) A tiling microarray expression analysis of rice chromosome 4 suggests a chromosome-level regulation of transcription. Plant Cell 17:1641–1657

Juretic N, Hoen DR, Huynh ML, Harrison PM, Bureau TE (2005) The evolutionary fate of MULE-mediated duplications of host gene fragments in rice. Genome Res 15:1292–1297

Kim C, Tang H, Paterson AH (2009) Duplication and divergence of grass genomes: integrating the chloridoids. Trop Plant Biol 2:51–62

Kim C, Zhang D, Auckland SA, Rainville LK, Jakob K, Kronmiller B, Sacks EJ, Deuter M, Paterson AH (2012) SSR-based genetic maps of Miscanthus sinensis and M. sacchariflorus,and their comparison to sorghum. TAG Theoretical and applied genetics 124:1325–1338

Kishimoto N, Higo H, Abe K, Arai S, Saito A, Higo K (1994) Identification of the duplicated segments in rice chromosomes 1 and 5 by linkage analysis of cDNA markers of known functions. Theor Appl Genet 88:722–726

Lande R (1975) Maintenance of genetic-variability by mutation in a polygenic character with linked loci. Genet Res 26:221–235

Lawrence WJC (1931) The secondary association of chromosomes. Cytologia 2:352–384

Lin Y-R (1998) Genetic analysis and progress in chromosome walking to the sorghum photoperiodic gene, Ma1. Department of Soil and Crop Science, Texas A&M, College Station, TX

Liu H, Sachidanandam R, Stein L (2001) Comparative genomics between rice and Arabidopsis shows scant collinearity in gene order. Genome Res 11:2020–2026

Ming R, Liu SC, Lin YR, da Silva J, Wilson W, Braga D, van Deynze A, Wenslaff TF, Wu KK, Moore PH, Burnquist W, Sorrells ME, Irvine JE, Paterson AH (1998) Detailed alignment of Saccharum and Sorghum chromosomes: comparative organization of closely related diploid and polyploid genomes. Genetics 150:1663–1682

Mondragon-Palomino M, Gaut BS (2005) Gene conversion and the evolution of three leucine-rich repeat gene families in *Arabidopsis thaliana*. Mol Biol Evol 22:2444–2456

Muller HJ (1964) The relation of recombination to mutational advance. Mutat Res 1:2–9

Nagamura Y, Inoue T, Antonio B, Shimano T, Kajiya H, Shomura A, Lin S, Kuboki Y, Harushima Y, Kurata N, Minobe Y, Yano M, Sasaki T (1995) Conservation of duplicated segments between rice chromosomes 11 and 12. Breed Sci 45:373–376

Ohta T (1984) Some models of gene conversion for treating the evolution of multigene families. Genetics 106:517–528

Paterson AH, Lin YR, Li ZK, Schertz KF, Doebley JF, Pinson SRM, Liu SC, Stansel JW, Irvine JE (1995a) Convergent domestication of cereal crops by independent mutations at corresponding genetic-loci. Science 269:1714–1718

Paterson AH, Schertz KF, Lin YR, Liu SC, Chang YL (1995b) The weediness of wild plants – molecular analysis of genes influencing dispersal and persistence of Johnsongrass, *Sorghum halepense* (L) Pers. Proc Natl Acad Sci USA 92:6127–6131

Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc Natl Acad Sci USA 101:9903–9908

Paterson AH, Bowers JE, Vandepoele K, Van de Peer Y (2005) Ancient duplication of cereal genomes. New Phytol 165:658–661

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Lyons E, Maher C, Narechania A, Penning B, Zhang L, Carpita NC, Freeling M, Jingle AR, Hash CT, Keller B, Klein PE, Kresovich S, McCann MC,

Ming R, Peterson DG, Ware D, Westhoff P, Mayer KFX, Messing J, Rokhsar DS (2009a) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboobur R, Ware D, Westhoff P, Mayer KF, Messing J, Rokhsar DS (2009b) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556

Petes TD, Symington LS (1991) The molecular and cellular biology of the yeast *Saccharomyces*: genome dynamics, protein synthesis and energetics. In: Broach JJ, Pringle J (eds) Recombination in yeast. Cold Spring Harbor Press, Cold Spring Harbor, NY

Prasad V, Stromberg CAE, Alimohammadian H, Sahni A (2005) Dinosaur coprolites and the early evolution of grasses and grazers. Science 310:1177–1180

Price HJ, Dillon SL, Hodnett G, Rooney WL, Ross L, Johnston JS (2005) Genome evolution in the genus *Sorghum* (Poaceae). Ann Bot 95:219–227

Salse J, Bolot S, Throude M, Jouffe V, Piegu B, Quraishi UM, Calcagno T, Cooke R, Delseny M, Feuillet C (2008) Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. Plant Cell 20:11–24

SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, MelakeBerhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL (1996) Nested retrotransposons in the intergenic regions of the maize genome. Science 274:765–768

Sawyer S (1989) Statistical tests for detecting gene conversion. Mol Biol Evol 6:526–538

Schmutz J, Cannon SB, Schlueter J, Ma JX, Mitros T, Nelson W, Hyten DL, Song QJ, Thelen JJ, Cheng JL, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu SQ, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du JC, Tian ZX, Zhu LC, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183

Simillion C, Vandepoele K, Van Montagu MCE, Zabeau M, Van de Peer Y (2002) The hidden duplication past of *Arabidopsis thaliana*. Proc Natl Acad Sci USA 99:13627–13632

Sobral BWS, Braga DPV, Lahood ES, Keim P (1994) Phylogenetic analysis of chloroplast restriction enzyme site mutations in the Saccharinae Griseb Subtribe of the Andropogoneae Dumort Tribe. Theor Appl Genet 87:843–853

Spangler R, Zaitchik B, Russo E, Kellogg E (1999) Andropogoneae evolution and generic limits in *Sorghum* (Poaceae) using ndhF sequences. Syst Bot 24:267–281

Swanson WJ, Clark AG, Waldrip-Dail HM, Wolfner MF, Aquadro CF (2001a) Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in Drosophila. Proc Natl Acad Sci USA 98:7375–7379

Swanson WJ, Zhang ZH, Wolfner MF, Aquadro CF (2001b) Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. Proc Natl Acad Sci USA 98:2509–2514

Swigonova Z, Lai J, Ma J, Ramakrishna W, Llaca V, Bennetzen JL, Messing J (2004a) Close split of sorghum and maize genome progenitors. Genome Res 14:1916–1923

Swigonova Z, Lai JS, Ma JX, Ramakrishna W, Llaca M, Bennetzen JL, Messing J (2004b) On the tetraploid origin of the maize genome. Comp Funct Genom 5:281–284

Szostak JW, Wu R (1980) Unequal crossing over in the ribosomal DNA of *Saccharomyces cerevisiae*. Nature 284:426–430

Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH (2008a) Synteny and colinearity in plant genomes. Science 320:486–488

Tang H, Wang X, Bowers JE, Ming R, Alam M, Paterson AH (2008b) Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. Genome Res 18:1944–1954

Tang H, Bowers JE, Wang X, Paterson AH (2010) Angiosperm genome comparisons reveal early polyploidy in the monocot lineage. Proc Natl Acad Sci USA 107:472–477

Van de Peer Y (2004) Computational approaches to unveiling ancient genome duplications. Nat Rev Genet 5:752–763

Vandepoele K, Simillion C, Van de Peer Y (2002) Detecting the undetectable: uncovering duplicated segments in Arabidopsis by comparison with rice. Trends Genet 18:606–608

Vandepoele K, Simillion C, Van de Peer Y (2003) Evidence that rice and other cereals are ancient aneuploids. Plant Cell 15:2192–2202

Vision T, Brown D, Tanksley S (2000) The origins of genomic duplications in Arabidopsis. Science 290:2114–2117

Wang X, Tang H, Bowers JE, Feltus FA, Paterson AH (2007) Extensive concerted evolution of rice paralogs and the road to regaining independence. Genetics 177:1753–1763

Wang X, Tang H, Bowers JE, Paterson AH (2009) Comparative inference of illegitimate recombination between rice and sorghum duplicated genes produced by polyploidization. Genome Res 19:1026–1032

White ME, Crowther BI (2000) Gene conversions may obscure actin gene family relationships. J Mol Evol 50:170–174

Yang ZH, Nielsen R, Goldman N, Pedersen AMK (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431–449

Zhang Y, Xu GH, Guo XY, Fan LJ (2005) Two ancient rounds of polyploidy in rice genome. J Zhejiang Univ Sci B 6:87–90

Chapter 19
Comparative Genomic Analysis of C4 Photosynthesis Pathway Evolution in Grasses

Xiyin Wang and Andrew H. Paterson

Abstract C4 plants are among the most productive crops, in part due to the efficiency at which their distinctive photosynthetic pathway fixes carbon at high temperatures. Sorghum and maize are C4 plants with full genome sequences available, facilitating a whole-genome level exploration of C4 pathway formation by comparing their respective versions of these key photosynthetic enzyme genes to those in the C3 plants rice and *Brachypodium*. A reservoir of duplicated genes was previously hypothesized to be a prerequisite for the evolution of C4 photosynthesis from a C3 progenitor. Grasses have been affected both by a whole-genome duplication (WGD) and individual gene duplications, and we show each of these mechanisms to have contributed to evolution of C4 photosynthesis. Some C4 genes appear to have been recruited directly from WGD duplicates followed by neofunctionalization. Others, such as the sorghum and maize carbonic anhydrase (CA) genes, have been recursively affected by tandem duplication, and mutations in stop codons have produced distinct C4 CA genes having 1–3 functional units, implying an interesting type of new gene formation accompanied by adaptive evolution. Key C4 enzymes in sorghum and maize show evidence of adaptive evolution, though differing in the level and mode. Intriguingly, a phosphoenolpyruvate carboxylase (PEPC) gene has

X. Wang (✉)
Plant Genome Mapping Laboratory, Departments of Crop and Soil Science, Plant Biology, and Genetics, University of Georgia, 2091 South Milledge Ave, C67, Athens, GA 30602, USA

Center for Genomics and Computational Biology, School of Life Sciences, Hebei United University, Tangshan, Hebei 063009, China
e-mail: wang.xiyin@gmail.com

A.H. Paterson
Plant Genome Mapping Laboratory, Departments of Crop and Soil Science, Plant Biology, and Genetics, University of Georgia, 2091 South Milledge Ave, C67, Athens, GA 30602, USA

A.H. Paterson (ed.), *Genomics of the Saccharinae*, Plant Genetics and Genomics: Crops and Models 11, DOI 10.1007/978-1-4419-5947-8_19,

also been evolving rapidly in both rice and *Brachypodium* and shows evidence of adaptive evolution, though lacking key mutations that are characteristic of C4 genes. The heterogeneity of origins of C4 genes suggests that there may have been a long transition process before the eventual establishment of C4 photosynthesis.

Keywords C4 photosynthesis pathway • Grass • Comparative genomics • Gene duplication • Adaptive evolution • Sorghum • Maize • Rice • Brachypodium

1 Introduction to the C4 Photosynthetic Pathway

Many of the most productive crops in agriculture are C4 plants. Despite their multiple origins, they are all characterized by high rates of photosynthesis and efficient use of water and nitrogen. As a morphological and biochemical innovation (Hatch and Slack 1966), the C4 photosynthetic pathway is proposed to have been an adaptation to hot, dry environments or CO_2 deficiency (Cerling et al. 1997; Ehleringer and Bjorkman 1978; Hattersley 1983; Seemann et al. 1987). The C4 pathway is thought to have independently appeared at least 50 times during angiosperm evolution (Mulhaidat et al. 2007; Sage 2004). Multiple origins of the C4 pathway within some angiosperm families (Giussani et al. 2001; Pyankov et al. 2001) imply that its evolution may not be so complex, perhaps suggesting that there may have been some genetic pre-deposition in some C3 plants to C4 evolution (Sage 2004).

The high photosynthetic capacity of C4 plants is due to their unique mode of CO_2 assimilation, featuring strict compartmentation of photosynthetic enzymes into two distinct cell types, mesophyll and bundle sheath (illustrated in Fig. 19.1 with NADP-ME C4 type). First, CO_2 assimilation is carried out in mesophyll cells. The primary carboxylating enzyme, phosphoenolpyruvate carboxylase (PEPC), together with carbonic anhydrase (CA) which is crucial to facilitating rapid equilibrium between CO_2 and HCO_3^-, is responsible for the hydration and fixation of CO_2 to produce a C4 acid, oxaloacetate (OAA). In NADP-malic enzyme-type C4 species, OAA is then converted to another C4 acid, malate, catalyzed by malate dehydrogenase (MDH). Malate then diffuses into chloroplasts in the proximal bundle sheath cells, where CO_2 is released to yield pyruvate by the decarboxylating NADP-malic enzyme (NADP-ME). The released CO_2 concentrates around the secondary carboxylase, Rubisco, and is reassimilated by it through the Calvin cycle. Pyruvate is transferred back into mesophyll cells and catalyzed by pyruvate orthophosphate dikinase (PPDK) to regenerate the primary CO_2 acceptor, phosphoenolpyruvate (PEP). Phosphorylation of a conserved serine residue close to the N-terminal end of the polypeptide of PEPC is essential to its activity by reducing sensitivity to the feedback inhibitor malate and a catalyst named PEPC kinase (PPCK). C4 photosynthesis has more efficient carbon assimilation at high temperatures because its combination of morphological and biochemical features reduce photorespiration, a loss of CO_2 that occurs during C3 photosynthesis at high temperatures (Sheen 1999). PPDK regulatory protein (RP), a bifunctional serine/threonine kinase-phosphatase,

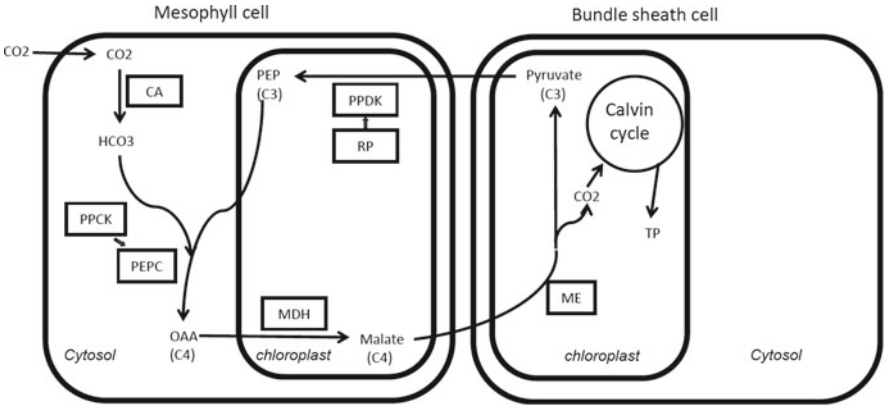

Fig. 19.1 NADP-ME type of C4 pathway in sorghum and maize. Abbreviations: *PEPC* phosphoenolpyruvate carboxylase, *PPCK* PEPC kinase, *NADP-MDH* malate dehydrogenase, *NADP-ME* NADP-malic enzyme, *PPDK* pyruvate, orthophosphate dikinase, *PPDK-RP* PPDK regulatory protein, *CA* carboxylating anhydrase, *TP* triosephosphate

catalyzes both the ADP-dependent inactivation and the Pi-dependent activation of PPDK (Burnell and Chastain 2006).

The evolution of a novel pathway is based on the creation of new genes, or functional changes in existing genes. Gene duplication is one of the principal mechanisms of the evolution of new genes. Genes encoding enzymes of the C4 cycle often belong to gene families having multiple copies. For example, maize and sorghum each have a single C4 PEPC gene and other non-C4 isoforms (Kawamura et al. 1992), whereas *Flaveria trinervia*, a C4 eudicot, has multiple copies of C4 PEPC genes (Poetsch et al. 1991). These findings led to the proposition that gene duplication, followed by functional innovation, was the genetic foundation for photosynthetic pathway transformation (Monson 2003).

All plant genomes, including grass genomes, have been enriched with duplicated genes derived from tandem duplications, single-gene duplications, and large-scale or whole-genome duplications (WGDs) (Blanc and Wolfe 2004b; Bowers et al. 2003; Wang et al. 2006; Yu et al. 2005a). A WGD occurred in a grass ancestor ~70 million years ago (mya), before the divergence of major cereal lineages including sorghum, maize, rice, wheat, and others (Paterson et al. 2004; Wang et al. 2005). A preliminary analysis of sorghum genome data suggested that duplicated genes from various sources have expanded the sizes of some families of C4-genes and their non-C4 isoforms (Paterson et al. 2009). However, different duplicated gene pairs often have divergent fates (Lynch and Conery 2003). While most duplicated genes are lost, gene retention in some functional groups sometimes produces large gene families in plants (Bowers et al. 2003; Paterson et al. 2004; Wang et al. 2005). Together with other lines of evidence, these have led to the interesting proposition of differential gene duplicability (He and Zhang 2005a; Liang and Li 2007), or duplication-resistance (Paterson et al. 2006), due to possible gene dosage imbalance,

which can be deleterious (Papp et al. 2003). Even when duplicated genes survive, there is rarely strong evidence supporting possible functional innovation (Nielsen et al. 2005).

Most C4 plants are grasses, and it has been inferred that C4 photosynthesis first arose in grasses during the Oligocene epoch (24–35 mya) (Christin et al. 2008; Vicentini et al. 2008). Sorghum and maize, thought to have diverged from a common ancestor ~12–15 mya (Paterson et al. 2009), are both in the Andropogoneae tribe, which is entirely composed of C4 plants (Giussani et al. 2001). Sorghum and maize, two representative species from the tribe Andropogoneae, composed of NADP-malic enzyme-type C4 plants, and their C3 relatives rice and *Brachypodium*, are among the early plant genomes to be sequenced (The International Brachypodium Initiative 2010; International Rice Genome Sequencing Project 2005; Paterson et al. 2009; Schnable et al. 2009; Yu et al. 2005b). The availability of these grass genome sequences using different types of photosynthesis provides a valuable opportunity to explore C4 pathway evolution. Here, we review a preliminary analysis (Wang et al. 2009) and extend upon it to compare C4 genes and their non-C4 isoforms in sorghum, maize, rice, and *Brachypodium*. Some *Saccharum* homologs were also involved if available. The aims of this study are to (1) investigate the role of gene duplication in the evolution of C4 enzyme genes; (2) investigate the role of adaptive evolution in C4 pathway formation; (3) investigate the long-standing hypothesis that a reservoir of duplicated genes has been a prerequisite of C4 pathway evolution (Monson 2003); and (4) investigate whether codon usage bias has contributed to C4 gene evolution, as previously suggested (Shenton et al. 2006). Our results will help to clarify the evolution of the C4 pathway and may better inform efforts to transform C3 plants, such as rice, to C4 photosynthesis (Sheehy et al. 2008).

2 Results

2.1 Evolution of PEPC Enzyme Genes

Grass PEPC enzyme genes form a small gene family (Cretin et al. 1990; Cretin et al. 1991). There are five plant-type and one bacteria-type PEPC (Sb03g008410 and Os01g0110700) (Sanchez and Cejudo 2003) gene isoforms in sorghum and rice, respectively, excepting two likely pseudogenized rice isoforms (Os01g0208800 and Os09g0315700) having only 217 and 70 codons. There is one C4

PEPC in both sorghum (Sb10g021330) and maize (Zm2g083841 on chromosome 9). One duplicated copy produced by the maize-specific WGD may have been lost. Previous characterization indicated that transcripts of the sorghum C4 enzyme are > 20 times more abundant in mesophyll than in bundle sheath cells (Wyrich et al. 1998). One Saccharum non-C4 isoform (M86661) has been identified so far (Christin et al. 2007).

By analysis of gene collinearity, we investigated how genome duplication has affected the PEPC gene families in these grasses. The PEPC gene in rice

which is most similar to the sorghum C4 PEPC is Os01g0208700, orthologous to *Brachypodium* gene Bradi3g06620, sharing ~73% amino acid identity. This similarity raised the possibility that the two genes are orthologous, though they are not in colinear locations. The outparalogs, homologs produced by WGD in the common ancestor of sorghum and rice, of the sorghum and maize C4 PEPC genes are at homoeologous locations. This indicates that one of the WGD-duplicates was recruited into the C4 pathway. The other grass PEPC genes form orthologous groups (Fig. 19.2). Whether or not the genes from different orthologous groups are outparalogs could not be supported by collinearity inference associated with the pan-cereal genome duplication.

Grass PEPC genes show high GC content, like many other grass genes, apparently as a result of changes after the monocot–dicot split but before the radiation of the grasses (Carels and Bernardi 2000). The evolution of C4 PEPC genes in sorghum and maize has been proposed to have been accompanied by GC elevation, resulting in codon usage bias (Lepiniec et al. 1993). We found that C4 PEPC genes do have higher GC content than other sorghum and maize PEPC genes, especially at the third codon sites (GC3). The sorghum and maize C4 PEPC genes have GC3 ~84%, significantly higher than other genes in both species. The suspected rice ortholog Os01g0208700 has even higher GC3 content, ~92%. In contrast, GC3 of all Arabidopsis PEPC genes are <43%. This shows that the higher GC content in the C4 PEPC genes may not be related to the evolution of C4 function, as discussed below.

Maximum likelihood analysis supports possible adaptive evolution of C4 PEPC genes. First, characterization of nonsynonymous nucleotide substitution rates (Ka) supports rapid evolution of the C4 genes and their rice ortholog. Under a free-parameter model, the C4 genes are among the fastest-evolving genes in the gene family. Second, the C4 genes may have been positively selected. The Ka/Ks ratio is nearly tenfold higher on the branch leading to the last common ancestor of the sorghum and maize C4 genes, than on other branches after the rice–sorghum split. Though the ratio is <1, we propose that the evident difference in Ka/Ks between C4 and non-C4 genes may indicate positive selection in the C4 genes for the following reasons: (1) The criterion Ka/Ks > 1 has been proposed to be unduly stringent to infer positive selection (Roth and Liberles 2006); (2) The maximum likelihood analysis is conservative as reported previously (Nielsen et al. 2005); (3) The similar slow evolutionary changes in all non-C4 genes in sorghum, maize, and rice (Fig. 19.1a) implies elevated rates in the C4 genes, rather than purifying selection in the non-C4 genes.

C4 PEPC genes show elevated and aggregated amino acid substitutions especially in function-specific regions, providing further evidence of adaptive evolution (Wang et al. 2009). Comparison to their outparalogs and their nearest outgroup sequence suggests that C4 PEPC genes have accumulated ~100 putative substitutions over their full length, far more than non-C4 genes. The substitutions are referred to as putative since we cannot rule out the possibility of parallel and reverse mutations. However, the extremely significant difference strongly supports divergent evolution of C4 and non-C4 PEPC genes. The amino acid substitutions are not uniformly distributed along the lengths of the C4 genes, but concentrated in the C-terminal half,

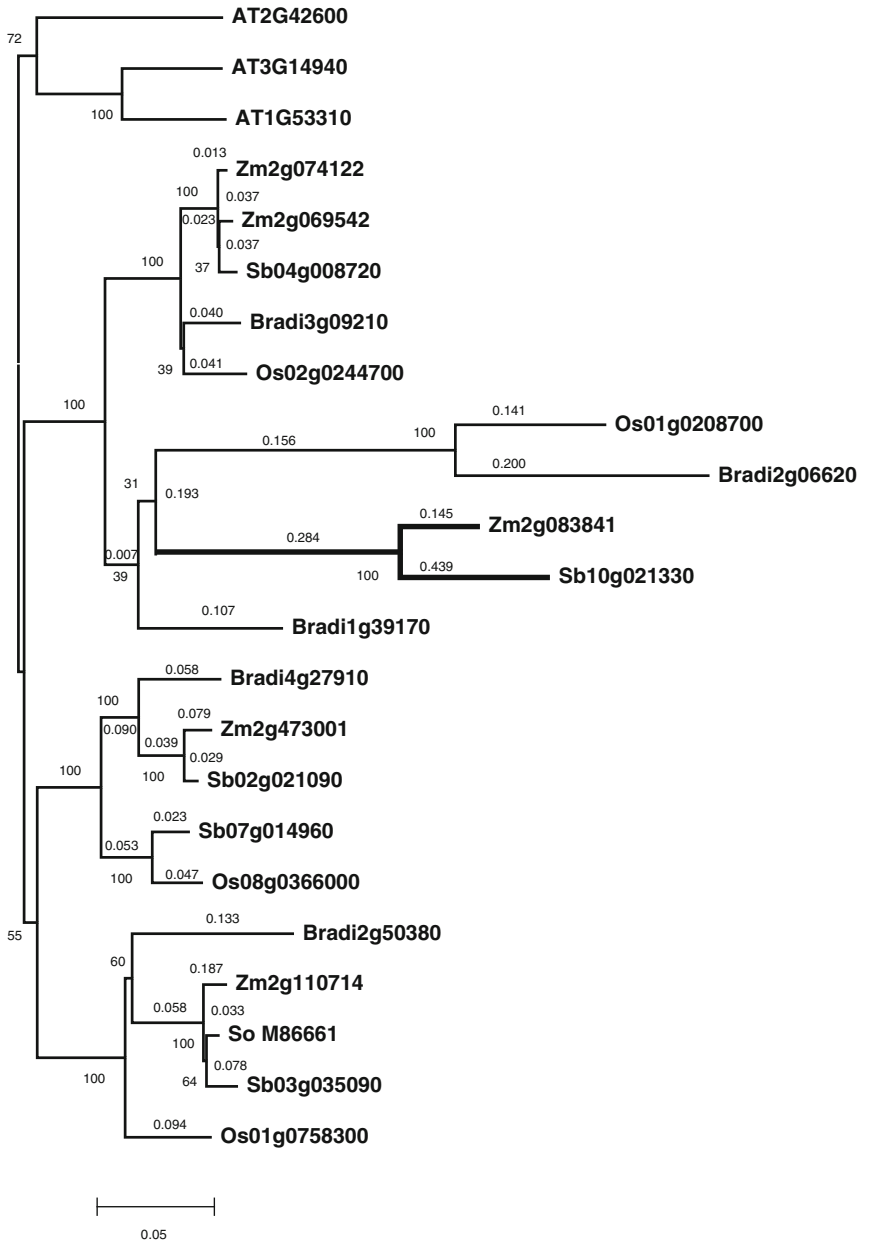

Fig. 19.2 Reconstructed phylogeny of PEPC enzyme genes and isoforms. *Thick branches* show C4 enzyme genes. Bootstrap values (integral numbers) and Ka/Ks ratios (numbers with fraction) are shown. The tree is produced by implementing the maximum likelihood approach by running PHYML with protein sequences using JTT model

including the critical mutation S780 (the serine at position 780 of maize C4 PEPC protein sequence). This is consistent with previous findings (Christin et al. 2007).

There is also evidence of adaptive evolution of possible C3 orthologs of the C4 PEPC genes. The rice gene Os01g0208700 and its *Brachypodium* ortholog Bradi2g06620 group together with the sorghum and maize C4 genes, and several outparalogs of the sorghum C4 gene form a sister group on the phylogenetic tree (Fig. 19.2). The pattern suggests that Os01g0208700 and Bradi2g06620 are orthologous to the sorghum C4 PEPC genes, implied by their high sequence similarity and common high GC content. The failure in collinearity inference could be explained by gene translocation in rice. Surprisingly, Os01g0208700 and Bradi2g06620 have also accumulated significantly more mutations than expected and have a relatively larger selection pressure than other non-C4 genes, implying that they may also be under adaptive selection as further discussed below.

2.2 Evolution of PPCK Enzyme Genes

PPCK gene families have been enriched by duplication events. We identified three PPCK 3 gene isoforms in rice, sorghum, and *Brachypodium*, respectively, and 4 in maize with two of these produced by a maize-specific duplication event. These genes form 3 groups on the phylogenetic tree, and are in expected colinear locations in each genome (Fig. 19.3). The sorghum and maize C4 PPCKs are encoded by Sb04g036570 and Zm2g178074, respectively. Their C4 nature is supported by evidence that their expression is light-induced and their transcripts are more abundant in mesophyll than bundle sheath cells (Shenton et al. 2006). In contrast, the expression of sorghum and maize non-C4 isoforms is not light- but cycloheximide-affected (Shenton et al. 2006). The outparalogs of the sorghum C4 gene and its rice ortholog were likely lost before the two species split, whereas the other 4 isoforms are outparalogs.

Maximum likelihood analysis and inference of aggregated amino acid substitutions found no evidence of adaptive selection during C4 PPCK gene evolution.

2.3 Evolution of MDH Enzyme Genes

There are two NADP-MDH enzyme genes in sorghum, the non-C4 gene Sb07g023910 and the C4 gene Sb07g023920, tandemly located as previously reported (Luchetta et al. 1991). They have only one homolog in rice (Os08g0562100), *Brachypodium* (Bradi3g12460), and maize (Zm2g129513 on chromosome 1), which are all at the expected colinear location. This suggests that the NADP-MDH WGD outparalog were lost before the sorghum–rice split, and that maize has lost additional copies since its lineage-specific WGD. Each of the sorghum tandem genes has an ortholog in Vetiveria and Saccharum, respectively (Rondeau et al. 2005), suggesting that the tandem duplication occurred before the divergence of sorghum and Vetiveria, but after the

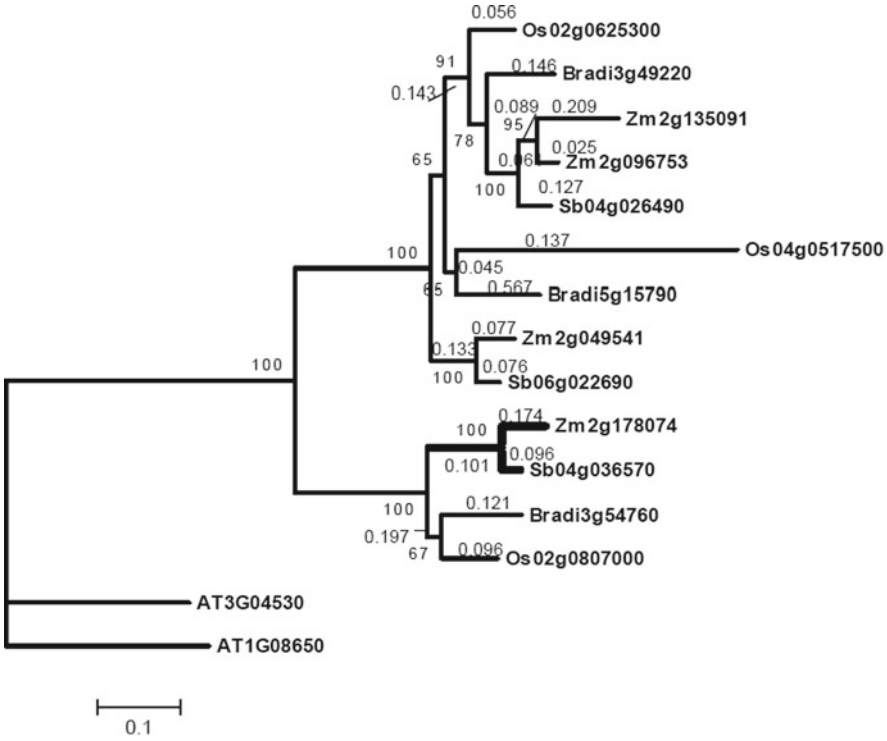

Fig. 19.3 Reconstructed phylogeny of PPCK enzyme genes and isoforms, produced by implementing the maximum likelihood approach by running PHYML with protein sequences using JTT model. *Thick branches* show C4 enzyme genes. Bootstrap values (integral numbers) and Ka/Ks ratios (numbers with fraction) are shown

sorghum–maize split, an inference further supported by gene tree analysis in that they are more similar to one another than to the single maize homolog (Fig. 19.4). There is also one C4 enzyme in *Saccharum* (AJ344432 in *Saccharum officinarum L.* and AJ416567 in *Saccharum spontaneum L).*

The C4 NADP-MDH gene shows an interesting mode of adaptive evolution. Clues of adaptive evolution include that: (1) the C4 NADP-MDH genes have accumulated more mutations than non-C4 genes; (2) The C4 gene cluster has a large Ka/Ks ratio (3.8), which is an indication of positive selection (Fig. 19.4), though not significantly supported by maximum likelihood analysis; (3) The C4 enzyme genes evolve faster than the non-C4 genes in sorghum and Saccharum. The sorghum C3 and C4 genes were likely to have been produced by an ancestral C4 gene through duplication. One of the duplicates may have lost its C4 function, for being not light-induced and only constitutively expressed (Luchetta et al. 1991).

The NADP-MDH genes are chloroplastic. A chloroplast transit peptide (cTP) having ~40 amino acids is identified in all the genes from grasses and Arabidopsis. This indicates that the cTP is ancestral in angiosperms. Non-chloroplastic

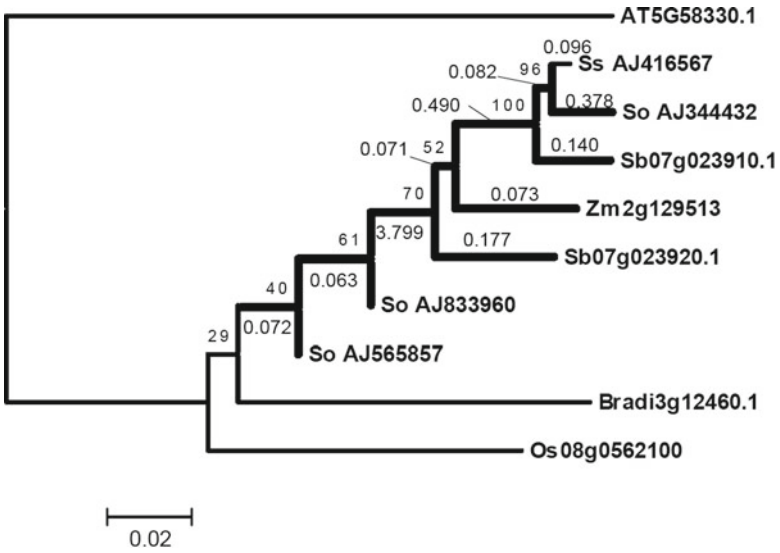

Fig. 19.4 Reconstructed phylogeny of NADP-MDH enzyme genes and isoforms, produced by implementing the maximum likelihood approach by running PHYML with protein sequences using JTT model. *Thick branches* show C4 enzyme genes. Bootstrap values (integral numbers) and Ka/Ks ratios (numbers with fraction) are shown

NADP-MDH genes identified in the sorghum genome share less than 40% protein sequence similarity with the chloroplastic ones.

2.4 NADP-ME Enzyme Genes

The NADP-ME gene family has been gradually expanding due to tandem duplication and the pan-cereal WGD. We identified 4–6 NADP-ME enzyme genes in the four grass genomes studied, respectively. The sorghum C4 gene is Sb03g003230, whose transcript is abundant in bundle sheath but not mesophyll cells (Wyrich et al. 1998). The maize C4 gene is Zm2g085019, at a collinear position on chromosome 3.

The C4 gene has a tandem duplicate in sorghum but not in maize (Fig. 19.5). The maize C4 gene and its closest paralog are both on chromosome 3 but not in proximal positions or colinear locations in any duplicated block. The sorghum tandem genes and their maize C4 ortholog share the same rice and *Brachypodium* orthologs at the expected colinear location, and their WGD duplicates can be found at expected colinear location in both species. The other sorghum and rice NADP-ME genes form 2 orthologous pairs, having also remained at the colinear locations predicted based on the pan-cereal duplication.

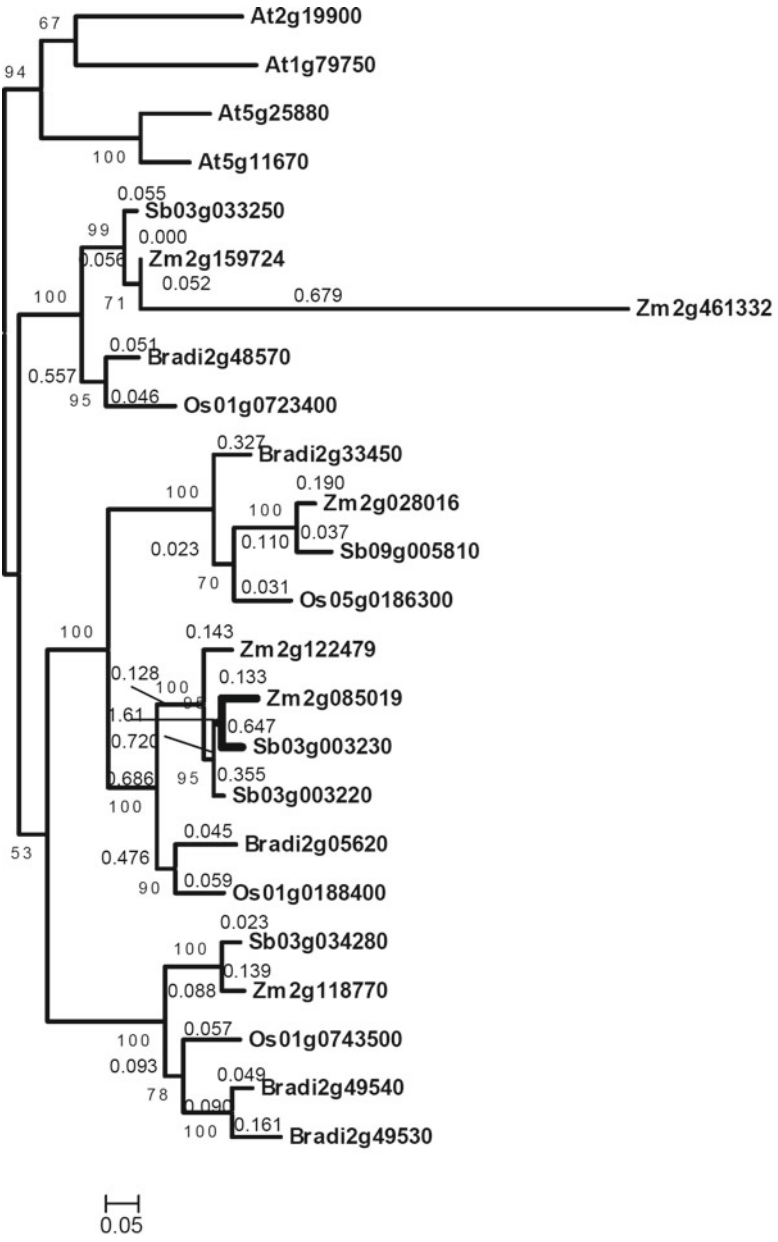

Fig. 19.5 Reconstructed phylogeny of NADP-ME enzyme genes and isoforms. *Thick branches* show C4 enzyme genes. Bootstrap values (integral numbers) and Ka/Ks ratios (numbers with fraction) are shown. The tree is produced by implementing the maximum likelihood approach by running PHYML with coding sequences using HKY model

Maximum likelihood analysis indicates that the sorghum and maize C4 NADP-ME genes are under positive selection. The branches leading to their two closest ancestral nodes have Ka/Ks ratios > 1. Moreover, the C4 genes have been significantly enriched by amino acid substitutions. The most affected regions in sorghum and maize overlap with one another, from residue 141 to residue 230 in sorghum, and from residue 69 to residue 181 in maize, respectively. This suggests that natural selection may have started to work in their common ancestor.

The C4 genes, their tandem paralog, and their rice ortholog all share a ~39-amino acid cTP that is not shared with their WGD paralogs in grasses, or with homologs in Arabidopsis.

2.5 PPDK Enzyme Genes

WGDs expanded the small PPDK gene family in grasses. The sorghum C4 PPDK gene (Sb09g019930) is identified based on its ~90% amino acid identity with the maize C4 gene (Zm2g011507 on chromosome 8). Its transcript is abundant in mesophyll rather than bundle sheath cells (Wyrich et al. 1998). The maize C4 gene has a duplicated copy (Zm2g097457 on chromosome 6), which is not involved in C4 photosynthesis, likely produced by the maize-specific WGD based on a small Ks value (0.34) and sharing collinearity. The C4 gene encodes both a C4 transcript and a cytosolic transcript, controlled by distinct upstream regulatory elements (Sheen 1991). The C4 transcript has a cTP encoded by an extra exon at a site upstream of the cytosolic gene (Glackin and Grula 1990). We found that the sorghum C4 PPDK gene shares high similarity to its maize counterpart along their respective full lengths, indicating their origin in a common maize-sorghum ancestor. We also involve one sugarcane (AF194026) and two Miscanthus C4 genes (AY262272 and AY262273) in the present analysis. Their rice and *Brachypodium* orthologs (Os05g0405000, Bradi2g25740) can be inferred based both on reconstruction of gene phylogeny (Fig. 19.6) and on gene collinearity. The other rice and sorghum isoforms are orthologous to one another. Whether the C4 and non-C4 isoforms are outparalogs produced by the WGD could not be determined by gene collinearity inference for possible gene translocations. However, synonymous nucleotide substitution rates and tree construction support that the rice and sorghum paralogs were produced before the two species diverged, and approximately at the pan-cereal WGD time.

The C4 PPDK genes may have been under adaptive evolution. While maximum likelihood analysis did not find evidence of adaptive evolution of C4 PPDK genes (Fig. 19.6), inference of amino acid substitution indicates that the C4 genes have accumulated significantly or nearly significantly more substitutions than their rice orthologs, particularly in the region from residue ~207 to residue ~620.

All the characterized isoform sequences from grasses and Arabidopsis share a ~20-amino acid cTP, suggesting its origin before the monocot–dicot split.

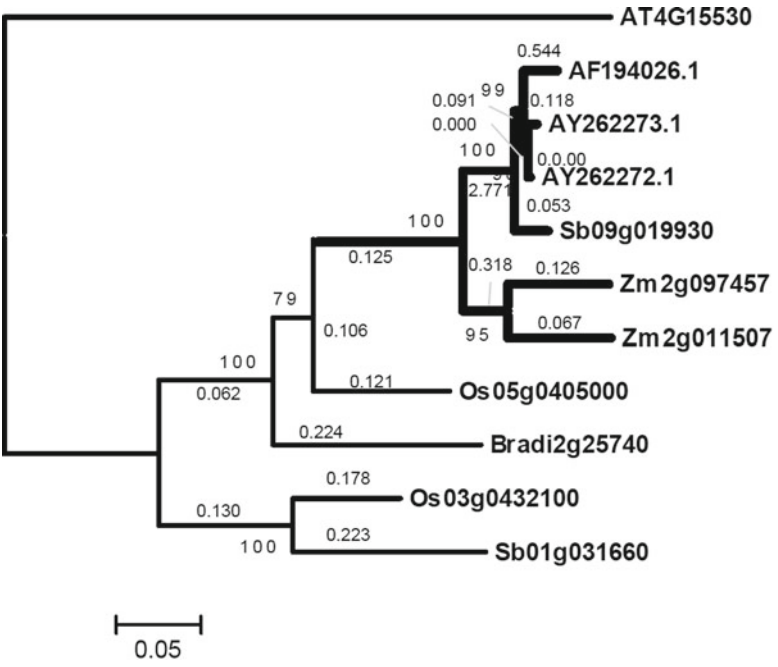

Fig. 19.6 Reconstructed phylogeny of PPDK enzyme genes and isoforms. *Thick branches* show C4 enzyme genes. Bootstrap values (integral numbers) and Ka/Ks ratios (numbers with fraction) are shown. The tree is produced by implementing the maximum likelihood approach by running PHYML with coding sequences using HKY model

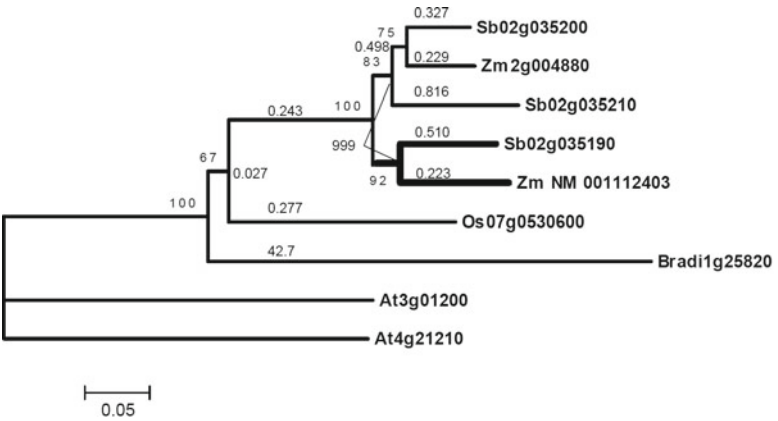

Fig. 19.7 Reconstructed phylogeny of PPDK-RP enzyme genes and isoforms, produced by implementing the maximum likelihood approach by running PHYML with protein sequences using JTT model, rooted using Arabidopsis homologs, and further edited by MEGA. *Thick branches* show C4 enzyme genes. Bootstrap values (integral numbers) and Ka/Ks ratios (numbers with fraction and those > 100) are shown

2.6 PPDK-RP Enzyme Genes

Tandem duplication contributed to the expansion of PPDK-RP genes Fig. 19.7, forming a small family. Using the maize PPDK-RP gene sequence as a query, we determined its possible sorghum ortholog, Sb02g035190, which has two tandem paralogs. Their rice ortholog, Os07g0530600, was identified in the anticipated colinear region. However, we failed to find their WGD outparalogs in both sorghum and rice, suggesting possible gene loss in their common ancestor. Moreover, the finding of only one copy in maize suggests loss of a second copy since the maize WGD.

Maximum likelihood analysis suggests that both lineages leading to the maize RP gene and its sorghum ortholog, and other isoforms have been under significant positive selection (Ka/Ks≫1), implying possible functional changes in both lineages. Compared to their rice ortholog, sorghum and maize RP genes have accumulated significantly more amino acid substitutions, providing supportive evidence for functional innovation.

2.7 CA Enzyme Genes

Tandem duplication has profoundly affected the evolution of CA genes. There are two types of CA enzymes, the alpha and beta types in grasses (Paterson et al. 2009), and C4 CA genes are the beta type (Tiwari et al. 2005). Our analysis indicates that there are four beta-type CA enzyme gene isoforms in sorghum, forming a tandem gene cluster with the same transcriptional orientation, on chromosome 3. Among these are two possible C4 genes (and Sb03g029180), which were shown to be highly expressed in mesophyll but not bundle sheath cells (Wyrich et al. 1998). The other two genes include one non-C4 gene (Sb03g029190) and one probable pseudogene (Sb03g029200) with only truncated coding sequence, a large DNA insertion in its second exon, and accumulated point mutations. The maize orthologs were duplicated during maize-specific WGD and preserved in the expected locations. There are three tandem rice genes and two tandem *Brachypodium* genes at the expected colinear location, indicating that gene family expansion had started before the divergence of the major cereal lineages but continuing in sorghum and maize after their split (See below for further information). The WGD outparalogs were not identified in either genome, implying possible gene loss after the WGD and before the rice–sorghum split.

The non-C4 gene, Sb03g029190, and its closest rice and *Brachypodium* collinear relatives on the phylogenetic tree have a normal structure, and the pseudogene, Sb03g029200, has a truncated structure.

The tandem duplication and gene fusion are shared by sorghum and maize, and maize may have an additional duplication. The coding sequence of a maize CA enzyme gene downloaded from GenBank has three functional units (Fig. 19.8), implying further DNA sequence duplication and gene fusion. Mutation of stop codons was also found in the leading gene sequences. Rice, *Brachypodium*, and *Arabidopsis* genes have only one functional unit preceded by a cTP.

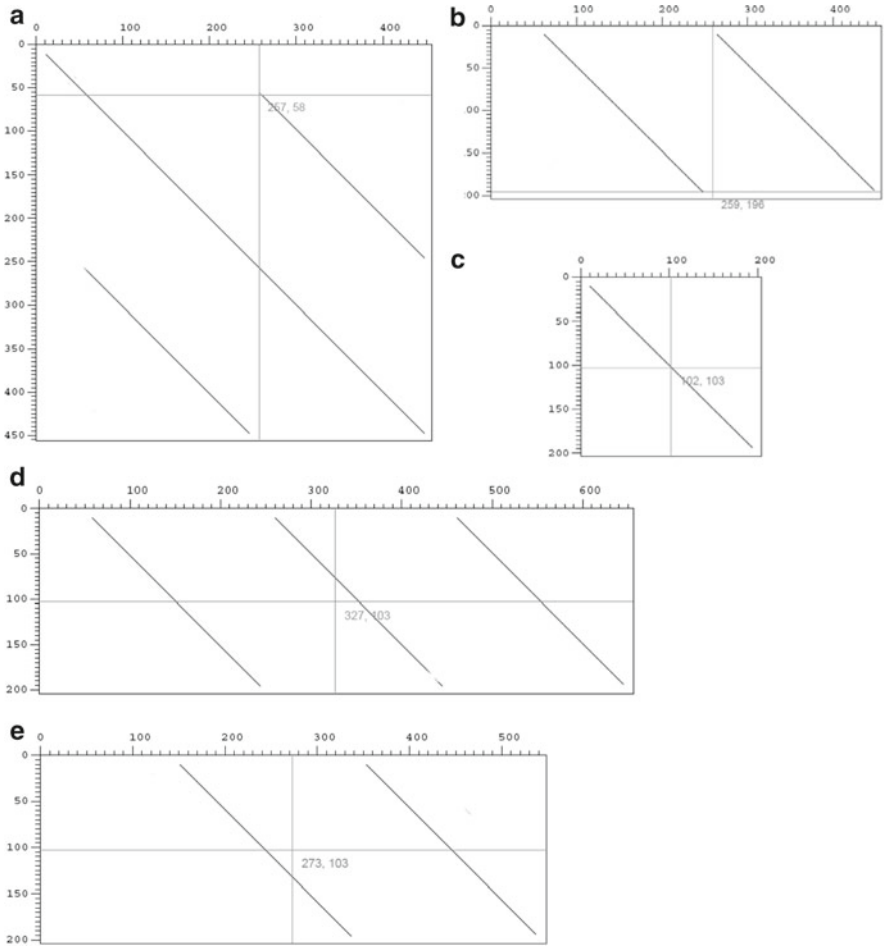

Fig. 19.8 Dotplots between sorghum and maize CA enzyme protein sequences. (**a**) Self-comparison of protein sequence of Sb03g029170. (**b**) Sb03g029170 (*horizontal*) and Sb03g029180 (*vertical*); (**c**) Sb03g029190 (*horizontal*) and Sb03g029180 (*vertical*); (**d**) Maize U08403 (*horizontal*) and Sb03g029180 (*vertical*); (**e**) Maize U08401 (*horizontal*) and Sb03g029180 (*vertical*)

To clarify the evolution of CA genes, we performed a phylogenetic analysis of the functional units (Fig. 19.9). The first functional units from sorghum and maize genes are grouped together, the second and third units and that of Sb03g029180 were in another group, and the rice gene and one sorghum gene Sb03g029190 were outgroups. This suggests the origin of the extra functional units to be after the Panicoideae-Ehrhartoideae divergence but before sorghum–maize divergence, and continuing in the maize lineage. A possible evolutionary process in sorghum is illustrated here (Fig. 19.10).

A gene tree of functional units suggested that C4 CA genes may have been affected by positive selection. According to the free-parameter model of the

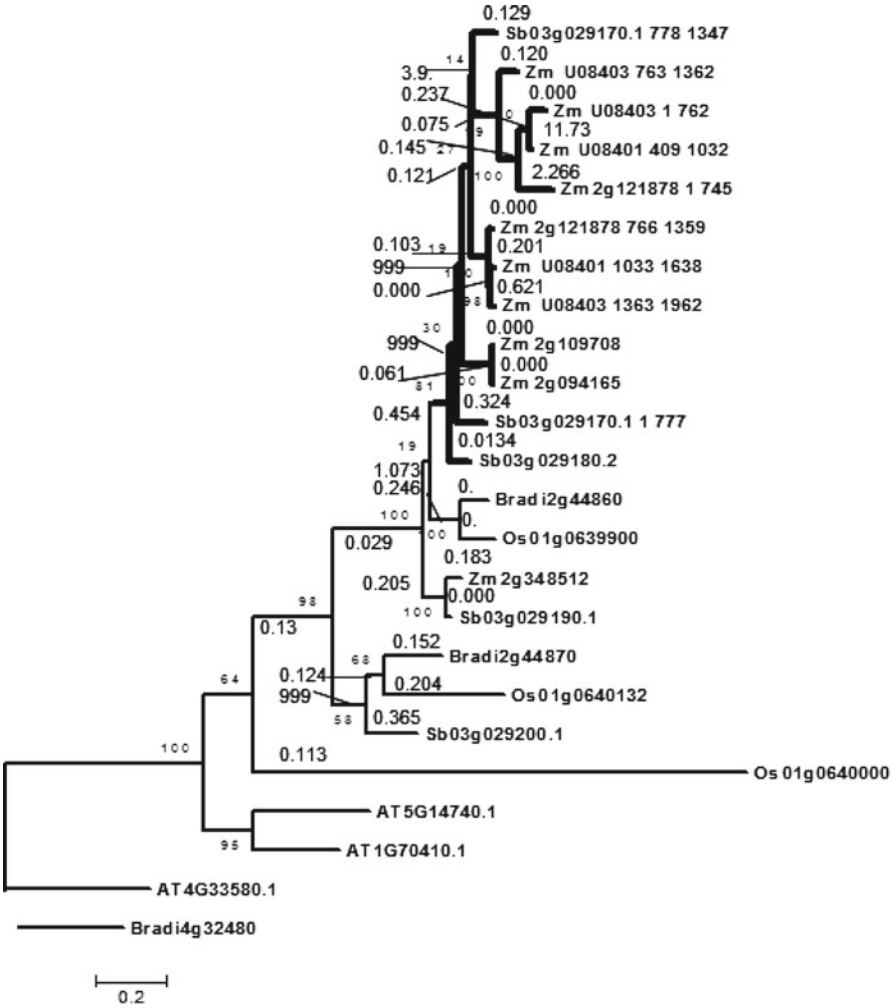

Fig. 19.9 Reconstructed phylogeny of functional units in CA enzyme genes and isoforms, produced by implementing the maximum likelihood approach by running PHYML with coding sequences using HKY model. For proteins having multiple functional units, these units are shown with gene Ids with starting and ending positions of the units. For example, "Sb03g029170.1 778 1347" shows the second functional unit (from 778th to 1347th amino acid) of the protein. *Thick branches* show C4 enzyme genes. Sorghum and maize C4 genes are shown with their functional repetitive units, indicated with starting and ending positions on the respective CDS sequences. Bootstrap values (integral numbers) and Ka/Ks ratios (numbers with fraction and those >100) are shown

maximum likelihood approach, we found that the two functional unit groups revealed above may have experienced positive selection, in that Ka/Ks > 1 (Fig. 19.9), though this possibility is not significantly supported by statistical tests or by amino acid substitution analysis.

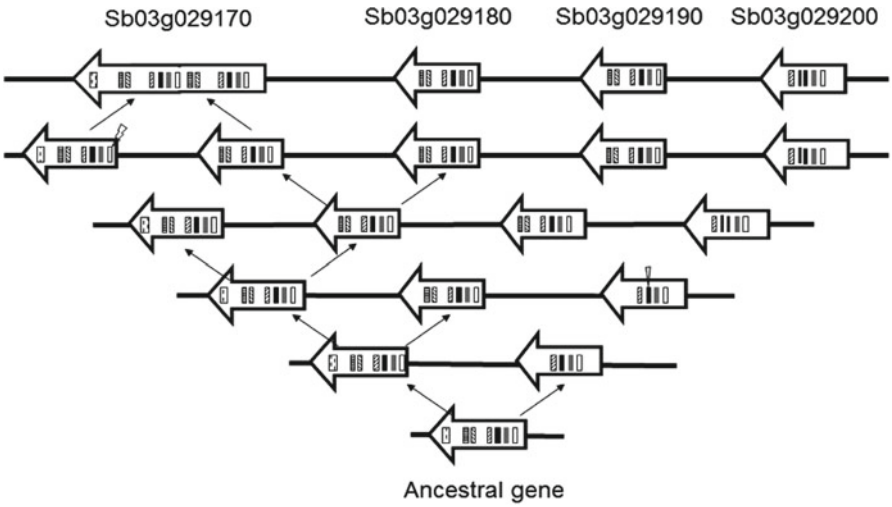

Fig. 19.10 Tandem duplication and fusion of CA genes in sorghum. Postulated evolution of sorghum CA genes through four tandem duplication events and a gene fusion event is displayed. We show distribution and structures of CA genes, and their peptide-encoding exons, on sorghum chromosome 3. Genes are shown in internally *hollow arrows*, and exons are shown as *blocks* with various internal patterns contained in the *arrows*. Homologous exons are with the same internal pattern. A chloroplast transit peptide is shown with *broken lines* as internal pattern. A tandem duplication event is shown with *two black arrows* pointing in divergent directions, and a gene fusion event shown with *two black arrows* pointing in convergent directions. A gene produced by fusion of two neighboring genes is shown in a bipartite structure. A stop codon mutation is shown with a *lightning-bolt mark*, and an exon-splitting event shown with a *narrow triangle*

3 Discussion

3.1 Gene Duplication and C4 Pathway Evolution

The evolution of the C4 pathway required the availability of gene families with multiple members, in which modification of both expression patterns and functional domains led to a new adaptive phenotype. An intuitive idea is that genetic novelty formation is simplified by exploiting available "construction bricks," and the pathway genes that we are aware of were either "subverted" from existing functions or were created through modification of existing genes. Three mechanisms of new gene formation have been proposed (Wolfe and Li 2003): duplication of preexisting genes followed by neofunctionalization, creation of mosaic genes from parts of other genes, and de novo invention of genes from DNA sequences.

Duplicated genes have long been suggested to contribute to the evolution of new biological functions. As early as 1932, Haldane suggested that gene duplication might have contributed new genetic materials because they create initially identical copies of

genes, which could be altered later to produce new genes without disadvantage to the organism (Haldane 1932). Ohno (1967) proposed that gene duplication played an essential role in evolution, pointed out the importance that WGD might have had on speciation, and hypothesized that at least one WGD event facilitated the evolution of vertebrates (Ohno 1970). This hypothesis has been supported by evidence from various gene families, and from the whole genome sequences of several metazoans (Meyer and Van de Peer 2005; Steinke et al. 2006). Plant genomes have experienced recurring WGDs (Bowers et al. 2003; Chapman et al. 2006; Jaillon et al. 2007; Soltis 2005; Tang et al. 2008a), and perhaps all angiosperms are ancient polyploids (Soltis 2005). These polyploidy events contribute to the creation of important developmental and regulatory genes (Blanc and Wolfe 2004a; Freeling and Thomas 2006; Maere et al. 2005; Seoighe and Gehring 2004), and may have played an important role in the origin and diversification of the angiosperms (De Bodt et al. 2005). About 20 mya before the divergence of the major groups of Poaceae grasses (Paterson et al. 2004; Wang et al. 2005), an ancestor experienced a WGD, preceded by more ancient duplication events (Salse et al. 2008; Tang et al. 2010; Wang et al. 2006). It is tempting to link this WGD to the evolutionary success of grasses, including more than 10,000 species and covering about 20% of the earth's land surface (Shantz 1954), though this link has not been adequately justified.

Gene duplication has been related to the evolution of the C4 pathway, based on the finding that C4 enzyme genes are usually from families having multiple copies (Monson 2003). Consequently, an ability to create and maintain large numbers of duplicated genes has been supposed to be one precondition for certain taxa to develop C4 photosynthesis (Monson 2003; Sage 2004). It was even suggested that evolution of the C4 pathway is largely a story of gene duplication while plants are still in the ancestral C3 state (Monson 2003).

Different genes in the C4 pathway were affected in different ways and at different times by gene duplication. The ~70-mya pan-cereal WGD presumably enriched the reservoir of all genes at least initially; however, only a subset of the duplicated genes were preserved. For example, in sorghum, both duplicated copies were preserved for PEPC and NADP-ME genes, and one of the copies produced by WGD was later recruited into the C4 pathway. This finding highlights the contribution of WGD to C4 photosynthesis evolution. However, for NADP-MDH, CA, PPDK-RP, and PPCK enzyme genes, one of the WGD duplicates was probably lost. In two of these cases (CA, PPDK-RP) and also for NADP-ME, tandem gains of new genes after the sorghum–rice divergence appear to have preceded C4 evolution. This seems to suggest that availability of the pan-cereal duplicated copies ~70-mya was not by itself sufficient to initiate C4 evolution, although it is not clear whether what was lacking was genetic (a part of the machinery) or environmental (a sufficiently strong selective advantage to drive the transition). Tandem duplication of C4 NADP-MDH following the sorghum–maize divergence does not appear to have been essential to C4 evolution; indeed, one of the tandem genes appears to have lost C4 specificity.

3.2 Adaptive Evolution of C4 Genes

After duplication, there is evidence that some C4 genes were under adaptive evolution; however, selection pressures and evolutionary modes have varied. Both maximum likelihood inference and patterns of aggregated amino acid differences indicate that the C4 NADP-ME and PPDK-RP enzyme genes have been under strong selective pressure. Maximum likelihood inference also implies that CA C4 enzymes are under positive selection, while aggregated amino acid differences indicate that C4 PEPC and PPDK genes may have been under adaptive evolution. The sorghum C4 genes of PPCK and NADP-MDH enzymes have also accumulated more substitutions than their rice orthologs, though the difference is not statistically significant. Compared to their rice orthologs, PEPC and NADP-MDH C4 genes evolve at a faster rate, providing further evidence of adaptation.

In some cases (NADP-ME, CA, PEPC) evidence from C4 plants supports adaptive evolution of the C4 gene family members only—the non-C4 homologs in C4 plants show no evidence of adaptive evolution, although the PEPC gene does show evidence in both rice and *Brachypodium* (C3). Further, the strongest evidence of adaptive evolution is in the period when the C4 pathway is thought to have evolved, after the divergence of sorghum and rice, but before the divergence of sorghum and maize.

Adaptive evolution is further supported by evidence from gene expression patterns. PEPC, PPDK, and CA C4 genes are expressed ~20 times more in sorghum mesophyll than bundle sheath cells, while NADP-ME C4 genes are expressed much more in bundle sheath than mesophyll (Wyrich et al. 1998). The study of *Flaveria* intermediates shows that PEPC activity is increased ~40 times from C3 to full C4 species (Svensson et al. 2003), and NADP-ME activity is ~9 times higher in veins than mesophyll cells (Hibberd and Quick 2002).

During the process of adaptive evolution, a duplicated gene may gradually acquire a new function (neofunctionalization) or subdivide the functions of its progenitor with the other duplicated copy (subfunctionalization). Laboratory evolution experiments indicated that an evolving new gene can initially acquire increased fitness for a new function without losing its original function (Aharoni et al. 2005). This implies that a neofunctionalization process may begin with an initial subfunctionalization step, an implication that has been supported by theory (He and Zhang 2005b). It is unclear how long such a step may take. Here, with the C4 genes we found that it could take a long time. Previous publications found that both C4 and non-C4 sorghum NADP-MDH genes were expressed in green leaves, though the C4 gene had higher transcript accumulation (Luchetta et al. 1991; Rondeau et al. 2005). Together with maximum likelihood analysis involving more genes and different grasses, this finding indicated that C4 and non-C4 sorghum NADP-MDH genes, produced before sorghum–Vetiveria divergence, have experienced subfunctionalization (Rondeau et al. 2005). Sequence alignment here indicates beyond some substitution sites, that the sorghum non-C4 gene has been affected by three insertion and one deletion event(s) in its N-terminal coding sequence, suggesting functional innovation. Regardless of whether the process is subfunctionalization or

neofunctionalization, co-expression, albeit at divergent levels, of the two genes in green leaves suggests that the process may not yet be finished.

In addition to the possible sheltering effect of a duplicated copy when evolving genetic novelty, alternative splicing may further shelter functional changes. The maize PPDK gene (and probably also its sorghum ortholog) encoding C4 transcripts, also encodes cytosolic transcripts. If C4 transcripts confer a novel function, and non-C4 transcripts perform the original function, the genes may have retained the original function for millions of years while evolving a novel function. The state of bifunctionality will probably continue until possible genetic incompatibility, if any, accumulates to a point intolerable to fitness. Their rice homolog also has a dual promoter (Matsuoka 1995), implying that natural selection works on this available functional duality to evolve C4 function. Maize PPDK may not be the only case of such gene bifunctionality in the C4 pathway. As shown above, the sorghum CA gene, Sb03g029170, seems to have similar bifunctionality, encoding both C4 and non-C4 transcripts. Since its internally repeating structure may have been produced before sorghum–maize divergence, its maize homologs also share this bifunctionality, which may have existed for millions of years. These multiple cases in which alternative splicing may contribute a possible sheltering effect during evolution of new function by C4 genes imply that it (alternative splicing) may participate in other cases of evolution of genetic novelty.

We found that the sorghum and maize C4 PEPC genes are on a long branch, grouped together with their suspected rice ortholog, showing possible adaptive evolution based on both a high Ka/Ks ratio and elevated amino acid substitution. It is intriguing to ask whether possible adaptive evolution in the rice PEPC gene could be a foundation toward a new origin of the C4 pathway, or instead indicates non-C4 functional adaptation. Scrutiny of the rice PEPC sequence revealed only 2 of 12 amino acid substitutions that were previously inferred to be positively selected in C4 genes (Christin et al. 2007), and in particular it lacks the critical fixed mutation S780 which is shared by C4 PEPCs in other angiosperms (Svensson et al. 2003; Westhoff and Gowik 2004). This rice gene was classified into the ppc-B1 group (Christin et al. 2007) found only in the C3 grasses, suggesting that its adaptive evolution is not leading to C4 photosynthesis, but possibly to other functional novelty.

Adaptive evolution of PEPC may have some valuable implications for the discovery of multiple groups of PEPC genes defined previously (Christin et al. 2007). In some C4 grasses there are different groups of genes, ppc-B2 and ppc-C4, with one group of genes, ppc-B1 found only in C3, not C4 grasses. These findings show that in the C4 lineages after their divergence with the C3 lineages but perhaps prior to the evolution of the C4 pathway itself, there may have been further gene duplication(s), some of which may have contributed to the establishment of C4 photosynthesis.

3.3 A Novel Mode of New Gene Evolution

The CA enzyme genes display a novel mode of gene evolution and functional adaptation. As shown above, sorghum and maize C4 CA enzymes have one, two, or three

functional domains, produced through recursive duplications followed by a fusion process involving stop codon mutations in the leading domains. There have been at least four tandem duplication events in sorghum and its ancestral genomes. These tandem duplications started before sorghum–maize divergence, and possibly continued in the maize lineage. The recurrence of tandem duplications together with the subsequent merger process may have acted as a mode of adaptive evolution. The present CA enzymes are beta-type, which is a dimer having four zinc ions bound to the structure as active sites. Besides dimers, these enzymes can form tetramers, hexamers, or octamers (Tiwari et al. 2005), suggesting that the dimer may be a building block. Recruiting extra domains through tandem duplications may contribute to the formation of more complex structures, with more functional binding sites making them work more efficiently to stabilize the balance between CO_2 and HCO_3^-. The expanded gene structure of these sorghum and maize CA genes are unusual, since the cDNAs of *Urochloa paniculata* and *Flaveria bidentis*, both C4 plants, are normal in size (Moroney et al. 2001). Nonetheless, there is precedent for internal repetition of CA gene structure in red algae, *Porphyridium purpureum*, resulting in two sets of functional binding sites (Mitsuhashi et al. 2000). The recurrence of forming internally repeating structure in CA gene evolution supports our proposition of possible functional advantages.

We found that the sorghum and maize C4 CA genes share a cTP, which had not been expected since the enzymes were not found to be chloroplastically localized in C4 plants. In C3 plants, the most abundant CA activity is in the chloroplast stroma, while in C4 plants, the exact location of CA is less clear (Tiwari et al. 2005), but the most abundant CA activity is localized in the cytosol of mesophyll cells (Ku et al. 1996). The cTP of sorghum and maize C4 CA genes is similar to that of the *Arabidopsis* CA genes, suggesting its existence before monocot–dicot divergence. The preservation of a cTP in C4 genes for tens of millions of years cannot be explained as a mere relic but suggests possible multiple functionality. This inference is at least partially supported by the discovery of divergent functions implemented by two different transcripts produced by a single sorghum C4 gene, Sb03g029170. As shown above, the expression of the longer transcript is light-inducible, while that of the shorter one is not, indicating that the longer but not the shorter transcript may be involved in the C4 pathway.

3.4 A Long Transition Time from C3 to C4 Photosynthesis

Numerous evolutionary models have been proposed to explain the formation of the C4 pathway (Brown and Hattersley 1989; Edwards and Ku 1987; Rawsthorne 1992). In summary, seven significant phases are recognized toward establishment of C4 photosynthesis: (1) general preconditioning (e.g. gene duplication), (2) anatomical preconditioning (e.g. close veins), (3) enhancement of bundle sheath organelles, (4) establishment of photorespiratory CO_2 pump and transformation of glycine decarboxylase to bundle sheath cells, (5) enhancement of PEPC

activity, (6) integration, and (7) optimization (Sage 2004). Although many biological and anatomical changes are needed, multiple origins in tens of angiosperm families suggest that it is not difficult to evolve a novel C4 pathway. However, from an evolutionary viewpoint it is still interesting to ask whether a transition process of gene functional changes and/or enhancement is necessary before final establishment, and how long such a transition might take. There was a long time-lag between the initial decrease in CO_2 concentration at least 100 mya (Sage 2004), and the appearance of C4 plants about 24–35 mya (Christin et al. 2008; Vicentini et al. 2008) One proposed explanation for the time-lag was the lack of a sufficient reservoir of duplicated and neofunctionalized C3 genes to support C4 evolution (Monson 2003). All required genes are thought to have been available at once, among the duplicated copies produced by the WGD ~70 mya (Paterson et al. 2004; Wang et al. 2005), however availability of this "reservoir" of duplicated genes was not by itself sufficient to inaugurate C4 evolution. Only a subset of the 70-mya duplicated genes were actually recruited to C4 photosynthesis (PEPC and NADP-ME), with many others being lost from the common cereal ancestor, then reduplicating before C4 evolution could occur.

3.5 Differential Duplicability of C4 Genes and Their Non-C4 Isoforms

Gene fates following the pan-cereal genome duplication shows differential duplicability of C4 genes and their isoforms in grasses. Evidence from yeast indicates that gene redundancy tends to be preserved among some of the central proteins in the cellular interaction network (Kafri et al. 2008). Tens of plant genes were suggested to be duplication-resistant, and undergo convergent restoration to singleton status following several independent genome duplications (Paterson et al. 2006). Differential duplicability could be explained by gene dosage effects, organismal complexity, protein interaction centrality, and protein domain preference (Kafri et al. 2008; Liang and Li 2007; Papp et al. 2003; Paterson et al. 2006). Here, we have shown that some gene families, including PEPC, PPCK, CA, and NADP-ME genes, have been expanded by gene duplication, but not others such as PPDK genes. The families expanded by gene duplication tend to be multiple functional, such as PEPC and NADP-ME (Monson 2003). Different PEPC gene isoforms take on specific roles, including the regulation of ion balance, the production of amino-group acceptor molecules in symbiotic nitrogen fixation, and the initial fixation of C in C4 photosynthesis and Crassulacean acid metabolism (Gehring et al. 1998). NADP-ME catalyzes the oxidative breakdown of malate to form CO_2 and pyruvate in the C4 pathway. Its non-C4 functions include the provision of carbon skeletons from ammonia assimilation (Chopra et al. 2002) and reductant for wound-induced production of lignin and flavonoids (Casati et al. 1999; Maurino et al. 2001). CA genes are also prone to duplication, which may enhance their formation of more complex structures, as discussed above. Though further duplication is not required when a former C3 gene is finally

co-opted for C4 roles (Monson 2003), we found that the sorghum NADP-MDH C4 gene did experience a tandem duplication event, with only one duplicated copy preserving the C4 function through possible subfunctionalization (Rondeau et al. 2005). This implies that the sorghum NADP-MDH C4 gene itself may be duplication-resistant.

3.6 C4 Pathway and Codon Usage Bias

GC content elevation has resulted in codon usage bias (Carels and Bernardi 2000), which was hypothesized to have contributed to C4 adaptive evolution (Shenton et al. 2006). We have shown that, though the grass C4 genes and their isoforms always have a higher GC content than their Arabidopsis counterparts, there is often a non-C4 grass gene having higher GC content than the C4 one(s) (Wang et al. 2009). Thus, there is no clear evidence supporting co-variation between codon usage bias and C4 gene evolution. Base composition variation in grass genes remains a controversial topic involving transcription, translation, modification, and mutational bias (Shi et al. 2007; Wang et al. 2004; Wong et al. 2002).

3.7 Potential Contribution to Engineering New C4 Plants

A comprehensive characterization of the C4 pathway will help not only to understand how C4 photosynthesis evolves but may also better inform efforts to transform C3 plants into C4 plants. To perform such a transformation, one strategy is to incorporate the C4 pathway into C3 plants through recombinant DNA technology (Miyao 2003), transferring C4 genes into C3 plants and yielding high levels of C4 enzymes in desired locations (Fukayama et al. 2001; Ku et al. 1999). It is of great interest to transform rice, a staple food for more than half of the world population, to perform C4 function, as reviewed recently (Sheehy et al. 2008). However, combined overproduction of C4 enzymes (PEPC, PPDK, NADP-ME, and NADP-MDH) resulted in only slightly higher levels of CO_2 assimilation in transformed rice than nontransgenic rice (Taniguchi et al. 2008). This might indicate that not all components needed for C4 photosynthesis are known. Knowledge of the complete sorghum genome might help to identify such missing components. As also shown above, though often not statistically significant, the sorghum and maize C4 genes appear to have been under adaptive evolution in different modes and levels and show different duplicability. These findings may provide clues toward a successful transformation of rice to C4 photosynthesis. Alternatively, perhaps adaptation beyond what we have discerned in the PEPC gene has occurred in C3 lineages, which may mitigate the perceived weaknesses of C3 photosynthesis.

3.8 Summary

Both WGD and single-gene duplication have contributed to C4 pathway evolution in sorghum and maize. Some C4 genes (PEPC, PPCK, and NADP-ME) were recruited from duplicates produced by WGD. Sorghum and maize NADP-MDH, NADP-ME, and PPDK-RP C4 genes were affected by tandem duplication, with only one of the resulting copies involved in the C4 pathway. C4 genes show divergent duplicability. PEPC, NADP-ME, PPCK, and CA gene families were expanded by recursive duplication events, showing a duplication-philic nature, whereas NADP-MDH and PPDK are likely duplication-phobic. Further supporting evidence is that only one copy of NADP-MDH C4 gene duplicates preserves the C4 function.

We found evidence of adaptive evolution of most C4 genes studied. However, the mode and level of adaptation is divergent among C4 genes. Adaptive evolution is achieved though rapid mutations in DNA sequences, aggregated amino acid substitutions, and/or considerable increase of expression level in specific cells. Besides gene redundancy, we found that alternative splicing may have also sheltered the evolution of new function. Our analysis supports previous findings that maximum likelihood inference may be too conservative to find adaptive evolution. We found no evidence of co-variation between codon usage bias and C4 pathway development.

Grass CA genes have evolved in a specific pattern featuring recursive tandem duplication and neighboring gene fusion, which produced distinct isoforms having 1–3 functional units, showing a novel type of new gene formation. Two sorghum C4 CA genes have one and two functional units, while two characterized maize C4 CA genes have two and three functional units, respectively. The elongation of these genes by recruiting extra domains may contribute to the formation of more complex protein structures, as often observed in plants.

The hypothesis that a reservoir of duplicated genes in ancestral C3 plants was a prerequisite for C4 pathway development is only partially supported by present findings. Availability of the pan-cereal duplicated copies was not sufficient to inaugurate C4 evolution, since some were lost from the common cereal ancestor, then had to reduplicate in the sorghum–maize ancestor before C4 evolution could occur. However, C4 gene isoforms show quite divergent duplicability, and there has been quite a long time-lag between the gene duplication events and the appearance of C4 grasses. These findings suggest a long transition process, including different modes of functional innovation, before the eventual establishment of C4 photosynthesis.

4 Materials and Methods

4.1 Materials

Known C4 enzyme genes and their non-C4 isoforms in sorghum, maize, and Arabidopsis were downloaded from NCBI coreNucleotide database (http://www.ncbi.nlm.nih.gov/). Searching these known genes against sorghum, maize, rice, and *Brachypodium*

gene models by running BLAST (Altschul et al. 1990) (E_value $< 1 \times 10^{-5}$), we identified other enzyme genes in these organisms. By characterizing sequence similarity and constructing gene trees, possible C4 genes were determined. The enzymes revealed here are linked to expression data in sorghum reported previously (Wyrich et al. 1998) by comparing cDNA segments to gene sequences using BLAST.

4.2 Gene Collinearity Inference

The potential gene homology information defined by running BLAST were used as the input for MCscan (Tang et al. 2008b) to find homologous gene pairs in collinearity. The built-in scoring scheme for MCscan is min(−log10E_value, 50) for every matching gene pairs and −1 for each 10 Kb distance between anchors, and blocks having scores >300 were kept. The resulting syntenic chains were evaluated using a procedure by colinearScan (Wang et al. 2006) and E-value $< 1e^{-10}$ was used as a significance cutoff.

4.3 Gene Phylogeny Construction

We constructed phylogenetic trees using several approaches, including the neighbor-joining method, maximum likelihood method, minimal evolution method, and maximum parsimony method, implemented in NADP-MEGA (Tamura et al. 2007), PHYML (Guindon et al. 2005), and PHYLIP (Felsenstein 1992), on both DNA and protein sequences. While running PHYML, parameters were set as adopted previously (Christin et al. 2007). Bootstrap tests were performed with 100 repeats to produce percentage values, showing the stability of their topology. The trees mostly agreed with one another. When there was inconsistency, the tree most strongly supported by bootstrap values was adopted for the subsequent adaptive evolution inference. For example, the trees of CA functional units were inconsistent among methods, and the best-supported Neighbor-joining tree produced by protein sequences was adopted for further analysis.

4.4 Maximum Likelihood Inference of Adaptive Evolution

The tree constructed for the group of C4 enzyme genes and their non-C4 isoforms was used to perform further maximum likelihood analysis using program codeml in PAML (Yang and Nielsen 1998). To detect whether a specific C4 gene has been positively selected, we compared two types of competing models: a free-ratio model and a ratio-restriction model (Yang 1998). The free-ratio model assumes an independent Ka/Ks ratio for each branch, whereas the latter force the Ka/Ks ratio = 1

on the specific branch which is to be tested for positive selection, and for the other branches assumes independent ratios. Each model produces a likelihood, and the twofold difference between them follows a Chi-squared distribution with 1 degree of freedom.

4.5 Aggregated Amino Acid Substitution Analysis

We adopted a comparative genomic approach initially proposed by Wagner (Wagner 2007) to detect genes potentially under positive selection. The Wagner approach inferred positive selection pressure by detecting possible aggregation of amino acid replacement. Here, we inferred possible amino acid replacements by comparing the homologous enzyme gene pair containing a C4 gene and non-C4 gene (often a rice gene) against the aligned outgroup sequence. A replacement site was identified in the C4 sequence that differed from the corresponding sites in both the homologous sequence and the outgroup sequence, which are identical. We found the number of all replacements, m, grouping the C4 and non-C4 protein sequences. If the occurrences of these replacement sites were assumed to be Poisson distributed with a parameter λ, we can evaluate the chance of observing a specific number of consecutive replacement sites along a sequence. For simplicity in description, for each sequence we first defined a replacement position array: $x = (x_0, x_1, x_2, \cdots, x_t, x_{t+1})$, composed by all the positions x_i ($1 \le i \le$) of replacement sites and two ends of the sequence, i.e., $x_0 = 0$ and $x_{t+1} = n+1$, where n is the length of the alignment after purging gaps. Then we defined the replacement distance array, $d = (d_1, \cdots, d_{t+1})$, where $d_i = x_i - x_{i-1}$ ($1 \le i \le t+1$). The distance between two replacement sites $d_{i,k} = x_{i+k-1} - x_i$, where k is the number of the consecutive replacement sites in the corresponding sequence segment, follows a Pearson type III distribution following a probability density $\lambda(\lambda z)^{k-2} e^{-\lambda z} / \Gamma(k-1)$ (Wagner 1997), where $\Gamma(k-1) = (k-1)!$. We can estimate the Poisson parameter λ with $m/(2n)$. Supposing there are t_i replacement sites along the i-th sequence, obviously, we got $\sum_{i=1}^{2} t_i = m$. Therefore, we could estimate the probability $P(d_i,k)$ that k consecutive replacements in a distance between two replacement sites is smaller than the observed d_i by the following integration: $P(d_{i,k}) = \left(\lambda / \Gamma(k-1)\right) \int_0^{d_{i,k}} (\lambda z)^{k-2} e^{-\lambda z} dz$.

We evaluated the occurrence probability of observed distance between any two replacement sites, and the smallest probability was used to locate a region with the most aggregated replacements, which was taken to be significant after a Bonferroni correction by considering the number of all combinations of replacement sites $\binom{t+2}{2}$. The occurrence probability was calculated using R language (http://www.r-project.org/). If between two replacement sites, there were gaps in the aligned sequences, they were omitted to check for possible selection. We composed Perl scripts to implement the described approach.

4.6 Maize Homolog Characterization

Maize BACs were from the MaizeSequence database (http://www.maizesequence.org/). The maize genes were searched against the BAC sequences to reveal their chromosomal locations, local DNA structures, etc.

4.7 Chloroplastic Transit Peptide Inferrence

ChloroP1.1 (Emanuelsson et al. 1999) was used to predict the presence of chloroplast transit peptides (cTP) in the enzyme protein sequences and the location of potential cTP cleavage sites

4.8 Dotplotting

Dotplots between CA protein sequences were produced by running the public program DOTTER (Sonnhammer and Durbin 1995). The Dotplots were produced by matched strings from two protein sequences in comparison. The expected score per residue of the matched strings was set to be 40.

Acknowledgements We appreciate financial support from the US National Science Foundation (MCB-0450260), CGIAR Generation Challenge Program, and United Sorghum Checkoff Program.

References

Aharoni A, Gaidukov L, Khersonsky O, McQ GS, Roodveldt C, Tawfik DS (2005) The 'evolvability' of promiscuous protein functions. Nat Genet 37:73–76

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Blanc G, Wolfe KH (2004a) Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. Plant Cell 16:1679–1691

Blanc G, Wolfe KH (2004b) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell 16:1667–1678

Bowers JE, Chapman BA, Rong J, Paterson AH (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 422:433–438

Brown RH, Hattersley PW (1989) Leaf anatomy of C(3)-C(4) species as related to evolution of C(4) photosynthesis. Plant Physiol 91:1543–1550

Burnell JN, Chastain CJ (2006) Cloning and expression of maize-leaf pyruvate, Pi dikinase regulatory protein gene. Biochem Biophys Res Commun 345:675–680

Carels N, Bernardi G (2000) Two classes of genes in plants. Genetics 154:1819–1825

Casati P, Drincovich MF, Edwards GE, Andreo CS (1999) Malate metabolism by NADP-malic enzyme in plant defense. Photosynth Res 61:99–105

Cerling TE, Harris JM, MacFadden BJ, Leasey MG, Quade J, Eisenmann V, Ehleringer JR (1997) Global vegetation change through the Miocene/Pliocene boundary. Nature 389:153–158

Chapman BA, Bowers JE, Feltus FA, Paterson AH (2006) Buffering crucial functions by paleologous duplicated genes may impart cyclicality to angiosperm genome duplication. Proc Natl Acad Sci U S A 103:2730–2735

Chopra J, Kaur N, Gupta AK (2002) A comparative developmental pattern of enzymes of carbon metabolism and pentose phosphate pathway in mungbean and lentil nodules. Acta Physiol Plant 24:67–72

Christin PA, Salamin N, Savolainen V, Duvall MR, Besnard G (2007) C4 photosynthesis evolved in grasses via parallel adaptive genetic changes. Curr Biol 17:1241–1247

Christin PA, Besnard G, Samaritani E, Duvall MR, Hodkinson TR, Savolainen V, Salamin N (2008) Oligocene CO_2 decline promoted C4 photosynthesis in grasses. Curr Biol 18:37–43

Cretin C, Keryer E, Tagu D, Lepiniec L, Vidal J, Gadal P (1990) Complete cDNA sequence of sorghum phosphoenolpyruvate carboxylase involved in C4 photosynthesis. Nucleic Acids Res 18:658

Cretin C, Santi S, Keryer E, Lepiniec L, Tagu D, Vidal J, Gadal P (1991) The phosphoenolpyruvate carboxylase gene family of Sorghum: promoter structures, amino acid sequences and expression of genes. Gene 99:87–94

De Bodt S, Maere S, Van de Peer Y (2005) Genome duplication and the origin of angiosperms. Trends Ecol Evol 20:591–597

Edwards GE, Ku MSB (1987) Biochemistry of C3-C4 intermediates. In: Hatch MD, Boardman NK (eds) The biochemistry of plants. Academic, London, pp 275–325

Ehleringer JR, Bjorkman O (1978) A comparison of photosynthetic characteristics of encelia species possessing glabrous and pubescent leaves. Plant Physiol 62:185–190

Emanuelsson O, Nielsen H, von Heijne G (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Sci 8:978–984

Felsenstein J (1992) Phylogenies from restriction sites—a maximum-likelihood approach. Evolution 46:159–173

Freeling M, Thomas BC (2006) Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. Genome Res 16:805–814

Fukayama H, Tsuchida H, Agarie S, Nomura M, Onodera H, Ono K, Lee BH, Hirose S, Toki S, Ku MS, Makino A, Matsuoka M, Miyao M (2001) Significant accumulation of C(4)-specific pyruvate, orthophosphate dikinase in a C(3) plant, rice. Plant Physiol 127:1136–1146

Gehring HH, Heute V, Kluge M (1998) Toward a better knowledge of the molecular evolution of phosphoenolpyruvate carboxylase by comparison of partial cDNA sequences. J Mol Evol 46:107–114

Giussani LM, Cota-Sanchez JH, Zuloaga FO, Kellogg EA (2001) A molecular phylogeny of the grass subfamily Panicoideae (Poaceae) shows multiple origins of C4 photosynthesis. Am J Bot 88:1993–2012

Glackin CA, Grula JW (1990) Organ-specific transcripts of different size and abundance derive from the same pyruvate, orthophosphate dikinase gene in maize. Proc Natl Acad Sci U S A 87:3004–3008

Guindon S, Lethiec F, Duroux P, Gascuel O (2005) PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res 33:W557–W559

Haldane JBS (1932) The causes of evolution. Cornell University Press, Ithaca

Hatch MD, Slack CR (1966) Photosynthesis by sugar-cane leaves. A new carboxylation reaction and the pathway of sugar formation. Biochem J 101:103–111

Hattersley PG (1983) The distribution of C3 and C4 grasses in Australia in relation to climate. Oecologia 57:113–128

He X, Zhang J (2005a) Gene complexity and gene duplicability. Curr Biol 15:1016–1021

He XL, Zhang JZ (2005b) Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. Genetics 169:1157–1164

Hibberd JM, Quick WP (2002) Characteristics of C4 photosynthesis in stems and petioles of C3 flowering plants. Nature 415:451–454

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436:793–800

Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, Vezzi A, Legeai F, Hugueney P, Dasilva C, Horner D, Mica E, Jublot D, Poulain J, Bruyere C, Billault A, Segurens B, Gouyvenoux M, Ugarte E, Cattonaro F, Anthouard V, Vico V, Del Fabbro C, Alaux M, Di Gaspero G, Dumas V, Felice N, Paillard S, Juman I, Moroldo M, Scalabrin S, Canaguier A, Le Clainche I, Malacrida G, Durand E, Pesole G, Laucou V, Chatelet P, Merdinoglu D, Delledonne M, Pezzotti M, Lecharny A, Scarpelli C, Artiguenave F, Pe ME, Valle G, Morgante M, Caboche M, Adam-Blondon AF, Weissenbach J, Quetier F, Wincker P (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463–467

Kafri R, Dahan O, Levy J, Pilpel Y (2008) Preferential protection of protein interaction network hubs in yeast: evolved functionality of genetic redundancy. Proc Natl Acad Sci U S A 105: 1243–1248

Kawamura T, Shigesada K, Toh H, Okumura S, Yanagisawa S, Izui K (1992) Molecular evolution of phosphoenolpyruvate carboxylase for C4 photosynthesis in maize: comparison of its cDNA sequence with a newly isolated cDNA encoding an isozyme involved in the anaplerotic function. J Biochem 112:147–154

Ku MS, Kano-Murakami Y, Matsuoka M (1996) Evolution and expression of C4 photosynthesis genes. Plant Physiol 111:949–957

Ku MS, Agarie S, Nomura M, Fukayama H, Tsuchida H, Ono K, Hirose S, Toki S, Miyao M, Matsuoka M (1999) High-level expression of maize phosphoenolpyruvate carboxylase in transgenic rice plants. Nat Biotechnol 17:76–80

Lepiniec L, Keryer E, Philippe H, Gadal P, Cretin C (1993) Sorghum phosphoenolpyruvate carboxylase gene family: structure, function and molecular evolution. Plant Mol Biol 21: 487–502

Liang H, Li WH (2007) Gene essentiality, gene duplicability and protein connectivity in human and mouse. Trends Genet 23:375–378

Luchetta P, Cretin C, Gadal P (1991) Organization and expression of the two homologous genes encoding the NADP-malate dehydrogenase in Sorghum vulgare leaves. Mol Gen Genet 228:473–481

Lynch M, Conery JS (2003) The evolutionary demography of duplicate genes. J Struct Funct Genomics 3:35–44

Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y (2005) Modeling gene and genome duplications in eukaryotes. Proc Natl Acad Sci U S A 102: 5454–5459

Matsuoka M (1995) The gene for pyruvate, orthophosphate dikinase in C4 plants: structure, regulation and evolution. Plant Cell Physiol 36:937–943

Maurino VG, Saigo M, Andreo CS, Drincovich MF (2001) Non-photosynthetic 'malic enzyme' from maize: a constitutively expressed enzyme that responds to plant defence inducers. Plant Mol Biol 45:409–420

Meyer A, Van de Peer Y (2005) From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). Bioessays 27:937–945

Mitsuhashi S, Mizushima T, Yamashita E, Yamamoto M, Kumasaka T, Moriyama H, Ueki T, Miyachi S, Tsukihara T (2000) X-ray structure of beta-carbonic anhydrase from the red alga, Porphyridium purpureum, reveals a novel catalytic site for CO(2) hydration. J Biol Chem 275:5521–5526

Miyao M (2003) Molecular evolution and genetic engineering of C4 photosynthetic enzymes. J Exp Bot 54:179–189

Monson RK (2003) Gene duplication, neofunctionalization, and the evolution of C4 photosynthesis. Int J Plant Sci 164:S43–S54

Moroney JV, Bartlett SG, Samuelsson G (2001) Carbonic anhydrases in plants and algae. Plant Cell Environ 24:13

Mulhaidat R, Sage RF, Dengler NG (2007) Diversity of kranz anatomy and biochemistry in C4 eudicots. Am J Bot 94:20

Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, Hubisz MJ, Fledel-Alon A, Tanenbaum DM, Civello D, White TJ, Sninsky JJ, Adams MD, Cargill M (2005) A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol 3:e170

Ohno S (1967) Sex chromosomes and sex-linked genes. Springler-Verlag, Berlin

Ohno S (1970) Evolution by gene duplication. Springer, Berlin

Papp B, Pal C, Hurst LD (2003) Dosage sensitivity and the evolution of gene families in yeast. Nature 424:194–197

Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc Natl Acad Sci U S A 101:9903–9908

Paterson AH, Chapman BA, Kissinger JC, Bowers JE, Feltus FA, Estill JC (2006) Many gene and domain families have convergent fates following independent whole-genome duplication events in Arabidopsis, Oryza, Saccharomyces and Tetraodon. Trends Genet 22:597–602

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob-ur R, Ware D, Westhoff P, Mayer KF, Messing J, Rokhsar DS (2009) The Sorghum bicolor genome and the diversification of grasses. Nature 457:551–556

Poetsch W, Hermans J, Westhoff P (1991) Multiple cDNAs of phosphoenolpyruvate carboxylase in the C4 dicot Flaveria trinervia. FEBS Lett 292:133–136

Pyankov VI, Artyusheva EG, Edwards GE, Black CC Jr, Soltis PS (2001) Phylogenetic analysis of tribe Salsoleae (Chenopodiaceae) based on ribosomal ITS sequences: implications for the evolution of photosynthesis types. Am J Bot 88:1189–1198

Rawsthorne S (1992) Towards an understanding of C3-C4 photosynthesis. Essays Biochem 27:135–146

Rondeau P, Rouch C, Besnard G (2005) NADP-malate dehydrogenase gene evolution in Andropogoneae (Poaceae): gene duplication followed by sub-functionalization. Ann Bot (Lond) 96:1307–1314

Roth C, Liberles DA (2006) A systematic search for positive selection in higher plants (Embryophytes). BMC Plant Biol 6:12

Sage RF (2004) The evolution of C4 photosynthesis. New Phytol 161:341–370

Salse J, Bolot S, Throude M, Jouffe V, Piegu B, Quraishi UM, Calcagno T, Cooke R, Delseny M, Feuillet C (2008) Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. Plant Cell 20:11–24

Sanchez R, Cejudo FJ (2003) Identification and expression analysis of a gene encoding a bacterial-type phosphoenolpyruvate carboxylase from Arabidopsis and rice. Plant Physiol 132:949–957

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen W, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He R, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin J, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren L, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Van Buren P, Vaughn MW, Ying K, Yeh CT, Emrich SJ, Jia Y, Kalyanaraman A, Hsia AP, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia JM, Deragon JM, Estill JC, Fu Y, Jeddeloh JA, Han Y, Lee H, Li P, Lisch DR, Liu S, Liu Z, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang L, Yu Y, Zhang L, Zhou S, Zhu Q, Bennetzen JL, Dawe RK, Jiang J, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA,

Wilson RK (2009) The B73 maize genome: complexity, diversity, and dynamics. Science (New York, NY) 326:1112–1115
Seemann JR, Sharkey TD, Wang J, Osmond CB (1987) Environmental effects on photosynthesis, nitrogen-use efficiency, and metabolite pools in leaves of sun and shade plants. Plant Physiol 84:796–802
Seoighe C, Gehring C (2004) Genome duplication led to highly selective expansion of the *Arabidopsis thaliana* proteome. Trends Genet 20:461–464
Shantz HL (1954) The place of grasslands in the earth's cover of vegetation. Ecology 35:143–145
Sheehy JE, Mitchell PL, Hardy B (2008) Charting new pathways To C4 rice. World Scientific, Los Banos, Philippines
Sheen J (1991) Molecular mechanisms underlying the differential expression of maize pyruvate, orthophosphate dikinase genes. Plant Cell 3:225–245
Sheen J (1999) C4 gene expression. Annu Rev Plant Physiol Plant Mol Biol 50:187–217
Shenton M, Fontaine V, Hartwell J, Marsh JT, Jenkins GI, Nimmo HG (2006) Distinct patterns of control and expression amongst members of the PEP carboxylase kinase gene family in C4 plants. Plant J 48:45–53
Shi X, Wang X, Li Z, Zhu Q, Yang J, Ge S, Luo J (2007) Evidence that natural selection is the primary cause of the GC content variation in rice genes. J Integr Plant Biol 49:1393–1399
Soltis PS (2005) Ancient and recent polyploidy in angiosperms. New Phytol 166:5–8
Sonnhammer ELL, Durbin R (1995) A dot-matrix program with dynamic threshold control suitable for genomic DNA and protein sequence analysis. Gene 167:1–10
Steinke D, Hoegg S, Brinkmann H, Meyer A (2006) Three rounds (1R/2R/3R) of genome duplications and the evolution of the glycolytic pathway in vertebrates. BMC Biol 4:16
Svensson P, Blasing OE, Westhoff P (2003) Evolution of C4 phosphoenolpyruvate carboxylase. Arch Biochem Biophys 414:180–188
Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599
Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH (2008a) Synteny and collinearity in plant genomes. Science (New York, NY) 320:486–488
Tang HB, Wang XY, Bowers JE, Ming R, Alam M, Paterson AH (2008b) Unreveling ancient hexaploidy throught multiply-aligned angiosperm gene maps. Genome Res 18:1944–1954
Tang H, Bowers JE, Wang X, Paterson AH (2010) Angiosperm genome comparisons reveal early polyploidy in the monocot lineage. Proc Natl Acad Sci U S A 107:472–477
Taniguchi Y, Ohkawa H, Masumoto C, Fukuda T, Tamai T, Lee K, Sudoh S, Tsuchida H, Sasaki H, Fukayama H, Miyao M (2008) Overproduction of C4 photosynthetic enzymes in transgenic rice plants: an approach to introduce the C4-like photosynthetic pathway into rice. J Exp Bot 59:1799–1809
International Brachypodium Initiative (2010) Genome sequencing and analysis of the model grass Brachypodium distachyon. Nature 463:763–768
Tiwari A, Kumar P, Singh S, Ansari S (2005) Carbonic anhydrase in relation to higher plants. Photosynthetica 43:1–11
Vicentini A, Barber JC, Aliscioni SS, Ciussani LM, Kellogg EA (2008) The age of the grasses and clusters of origins of C4 photosynthesis. Glob Chang Biol 14:15
Wagner A (1997) A computational genomics approach to the identification of gene networks. Nucleic Acids Res 25:3594–3604
Wagner A (2007) Rapid detection of positive selection in genes and genomes through variation clusters. Genetics 176:2451–2463
Wang HC, Singer GA, Hickey DA (2004) Mutational bias affects protein evolution in flowering plants. Mol Biol Evol 21:90–96
Wang X, Shi X, Hao B, Ge S, Luo J (2005) Duplication and DNA segmental loss in the rice genome: implications for diploidization. New Phytol 165:937–946
Wang X, Shi X, Li Z, Zhu Q, Kong L, Tang W, Ge S, Luo J (2006) Statistical inference of chromosomal homology based on gene colinearity and applications to Arabidopsis and rice. BMC Bioinformatics 7:447

Wang X, Gowik U, Tang H, Bowers JE, Westhoff P, Paterson AH (2009) Comparative genomic analysis of C4 photosynthetic pathway evolution in grasses. Genome Biol 10:R68

Westhoff P, Gowik U (2004) Evolution of c4 phosphoenolpyruvate carboxylase. Genes and proteins: a case study with the genus Flaveria. Ann Bot (Lond) 93:13–23

Wolfe KH, Li WH (2003) Molecular evolution meets the genomics revolution. Nat Genet 33(Suppl):255–265

Wong GK, Wang J, Tao L, Tan J, Zhang J, Passey DA, Yu J (2002) Compositional gradients in Gramineae genes. Genome Res 12:851–856

Wyrich R, Dressen U, Brockmann S, Streubel M, Chang C, Qiang D, Paterson AH, Westhoff P (1998) The molecular basis of C4 photosynthesis in sorghum: isolation, characterization and RFLP mapping of mesophyll- and bundle-sheath-specific cDNAs obtained by differential screening. Plant Mol Biol 37:319–335

Yang Z (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol 15:568–573

Yang Z, Nielsen R (1998) Synonymous and nonsynonymous rate variation in nuclear genes of mammals. J Mol Evol 46:409–418

Yu J, Wang J, Lin W, Li S, Li H, Zhou J, Ni P, Dong W, Hu S, Zeng C, Zhang J, Zhang Y, Li R, Xu Z, Li S, Li X, Zheng H, Cong L, Lin L, Yin J, Geng J, Li G, Shi J, Liu J, Lv H, Li J, Wang J, Deng Y, Ran L, Shi X, Wang X, Wu Q, Li C, Ren X, Wang J, Wang X, Li D, Liu D, Zhang X, Ji Z, Zhao W, Sun Y, Zhang Z, Bao J, Han Y, Dong L, Ji J, Chen P, Wu S, Liu J, Xiao Y, Bu D, Tan J, Yang L, Ye C, Zhang J, Xu J, Zhou Y, Yu Y, Zhang B, Zhuang S, Wei H, Liu B, Lei M, Yu H, Li Y, Xu H, Wei S, He X, Fang L, Zhang Z, Zhang Y, Huang X, Su Z, Tong W, Li J, Tong Z, Li S, Ye J, Wang L, Fang L, Lei T, Chen C, Chen H, Xu Z, Li H, Huang H, Zhang F, Xu H, Li N, Zhao C, Li S, Dong L, Huang Y, Li L, Xi Y, Qi Q, Li W, Zhang B, Hu W, Zhang Y, Tian X, Jiao Y, Liang X, Jin J, Gao L, Zheng W, Hao B, Liu S, Wang W, Yuan L, Cao M, McDermott J, Samudrala R, Wang J, Wong GK-S, Yang H (2005a) The genomes of *Oryza sativa*: a history of duplications. PLoS Biol 3:e38

Yu J, Wang J, Lin W, Li SG, Li H, Zhou J, Ni PX, Dong W, Hu SN, Zeng CQ, Zhang JG, Zhang Y, Li RQ, Xu ZY, Li ST, Li XR, Zheng HK, Cong LJ, Lin L, Yin JN, Geng JN, Li GY, Shi JP, Liu J, Lv H, Li J, Deng YJ, Ran LH, Shi XL, Wang XY, Wu QF, Li CF, Ren XY, Wang JQ, Wang XL, Li DW, Liu DY, Zhang XW, Ji ZD, Zhao WM, Sun YQ, Zhang ZP, Bao JY, Han YJ, Dong LL, Ji J, Chen P, Wu SM, Liu JS, Xiao Y, Bu DB, Tan JL, Yang L, Ye C, Zhang JF, Xu JY, Zhou Y, Yu YP, Zhang B, Zhuang SL, Wei HB, Liu B, Lei M, Yu H, Li YZ, Xu H, Wei SL, He XM, Fang LJ, Zhang ZJ, Zhang YZ, Huang XG, Su ZX, Tong W, Li JH, Tong ZZ, Li SL, Ye J, Wang LS, Fang L, Lei TT, Chen C, Chen H, Xu Z, Li HH, Huang HY, Zhang F, Xu HY, Li N, Zhao CF, Dong LJ, Huang YQ, Li L, Xi Y, Qi QH, Li WJ, Hu W, Zhang YL, Tian XJ, Jiao YZ, Liang XH, Jin JA, Gao L, Zheng WM, Hao BL, Liu SQ, Wang W, Yuan LP, Cao ML, McDermott J, Samudrala R, Wong GKS, Yang HM (2005b) The genomes of Oryza sativa: a history of duplications. PLoS Biol 3:266–281

Chapter 20
Differentiation of Seed, Sugar, and Biomass-Producing Genotypes in *Saccharinae* Species

Seth C. Murray

Abstract Plant sinks including grain, stem sugar, leaf, and stem biomass as well as roots and rhizomes are the primary *Saccharinae* products of interest to humans. While plants evolved particular ratios of these sinks in the wild, humans have selected for altered ratios to fit production needs. Through further selection, specialty cultivars have recently been developed to maximize grain, sugar, biomass, or rhizome production as well as specialty carbohydrate and nutrient compositions. In sorghum and other C4 grass crops these specialized cultivars continue to undergo differentiation and could become, to some extent, genetically isolated populations. There has been limited research on the genetics and diversity of morphologically different crop ideotypes but it appears that inadequate recombination and selection has occurred to genetically isolate such genotypes yet. Importantly, recent results have shown that genetic tradeoffs among these various plant products may be avoidable, because photosynthesis is often sink, rather than source, limited. By combining and selecting multiple product types as optimized in specialty cultivars, harvestable products and energy may be further increased. Here, I review molecular genetic and phenotypic evidence of ideotype differentiation and discuss how harvestable energy might be maximized by carefully selecting for multiple products. There are a wide range of potential confounding factors, such as flowering time, that may alter conclusions and some possible solutions are suggested. If an ultimate goal is maximum usable energy production per unit area to satisfy growing global demand for food, fiber, fuel, and land, then breeding for only single products may sacrifice potential productivity.

Keywords Energy sinks • Genetic diversity • Genetic tradeoffs • NIRS • Phenotyping • Pleiotropy • Sweet sorghum • Yield

S.C. Murray (✉)
Department of Soil and Crop Sciences, Texas A&M University,
370 Olsen Blvd. 2474 TAMU, College Station, TX 77843-2474, USA
e-mail: sethmurray@tamu.edu

A.H. Paterson (ed.), *Genomics of the Saccharinae*, Plant Genetics and Genomics: Crops and Models 11, DOI 10.1007/978-1-4419-5947-8_20,

1 The Importance of Plant Energy Sinks

Wild plants experience constant natural selection to efficiently use and store energy from photosynthesis in different ways that maximize reproductive success. First, plants can reinvest this energy in products that allow the harvest of additional raw materials for energy production at a later date, such as roots to extract water and nutrients, or leaves to capture more sunlight and CO_2 for photosynthesis. Second, they can store this energy and additional nutrients in biomass, such as larger diameter or taller stalks that might allow them to better compete with already established neighbors (Givnish 1982). Third, they can invest in sexual propagules such as seeds that will insure survival in the next generation and through adverse environmental conditions. Fourth, they can partition energy into temporary nonstructural carbohydrate reserves such as stem sugar that can provide a buffer during short periods of stress (Blum 1996, 1998). Finally, they can invest jointly in both storage and nonsexual (clonal) reproduction in rhizomes (or in the case of dicots, tubers) that allows competition in the current and future generation even through adverse environments (Anderson et al. 1960; Williams and Ingber 1977). In the wild these are not mutually exclusive strategies, and instead quantitative variations in resource allocations can be identified both between and within species (Loehle 1987). Such variation in allocations can also be seen segregating within domesticated species. For instance, modern elite cultivars of wheat and rice were selected during the green revolution for dwarfing which improved grain harvest index (decreased biomass and increased grain) and have different photosynthetic allocations than tall landraces (Reynolds et al. 1999; Borlaug 2002). Though much work on carbohydrate resource and reproductive allocation has been conducted, carbon partitioning into various sinks and the underlying genetic variation remains poorly understood and varies among species, genotype, developmental stage, and environment (Marcelis and Heuvelink 2007).

Humans, through domestication and subsequent crop improvement, have selected from natural genetic diversity to maximize differentiation of and investment in specific plant photosynthate sinks for our own benefit. In few crops are there clearer examples of this differentiation than in sorghum, as described by Kresovich and Dahlberg in Chap. 2 of this book. Various types of sorghum have been selected to maximize energy invested in sexual reproduction and starch (grain sorghum), stem sugar storage (sweet sorghums), biomass (forage cultivars and energy cultivars), dual purpose crops (grain and forage cultivars in Africa and Asia), panicle branches (broomcorn), and rhizomes (weedy types inadvertently selected for agricultural systems by humans, perennial grain sorghum).

Based on shared evolutionary histories, the genetic variance observed and colinearity among genomes, it is likely that similar genetic potential exists to differentially improve multiple sink traits across the *Saccharinae* and allied species. So far, however, this has not been done in other taxa to the extent of sorghum. *Saccharum* has recently been differentially selected for both sweet (sugarcane) and biomass (energy cane) types. A close *Saccharinae* relative, maize (*Zea mays* L.), shows potential genetic variation for dedicated biomass, sweet stem, and perennial ideotypes, in addition to dedicated grain or silage cultivars (Dhugga 2007; Stewart 1878;

Shaver 1967; Murray personal observation). A few additional crops such as millet and small grains have also experienced divergent selection for sink ratios to develop grain, silage, and stover types.

Although grain and sugar have always been valued as commodities, aboveground and belowground cellulosic biomass are also increasingly valued. This interest in biomass centers around the potential use of cellulosic biofuel, ability to maintain/improve soil quality, the use of soil for carbon capture, and growing appreciation of the benefits of soil retention and energy provided by perennials. With advancements in technology many new plant products offer ecosystem services with value that is increasingly measurable and may warrant improvement through breeding.

Genomics of *Saccharinae* species and other crops have been primarily justified in the context of crop improvement. Here I review phenotypic differentiation in the *Saccharinae* and allied species and discuss how these findings can be combined with genomics to facilitate further breeding and improvement to meet both existing and anticipated food, feed, fiber, and fuel needs. Specifically, important questions include (1) Are there genetic tradeoffs in plant sink production and how do we appropriately evaluate these? (2) How do we improve plants to maximize energy production given emerging technology and a need to maintain whole plant homeostasis? (3) How do we value and optimize relationships between sink yield versus sink composition? (4) What is known about genetic diversity and differentiation of plant products that can assist in obtaining these goals?

2 A Question of Sink Tradeoffs?

Conventional dogma has suggested that there exist inevitable tradeoffs between plant sink products; in other words, for every joule that goes into grain 1 J less goes into stem sugar, biomass, or rhizomes. An unstated assumption in this thesis is that photosynthesis is source limited: in other words, at all times the plant can use or store the maximum amount of energy that can be produced. Recent evidence in many plants, especially C4 grasses, suggests that, given conventional access to limiting nutrients, source photosynthate production is often limited by sink mechanisms and transport (Paul and Foyer 2001; Reynolds et al. 2005; Dingkuhn et al. 2007). Lines of supporting evidence across the *Saccharinae* and allied species will be further described below. It is central to the concept of future crop improvement that the classical assumption of plant sink "inevitable tradeoffs" is in need of revision.

2.1 Sorghum

Sorghum has more differentiated special-use genotypes than most other crops, including grain-, forage-, sugar-, energy- (biomass), fiber-, and broom panicle-producing genotypes (Harlan and deWet 1972; Mullet et al. 2010). As Kresovich and Dahlberg (Chap. 2) discuss, it therefore makes an outstanding case study to evaluate the effects

of directional selection, multiple product types, and the genomics of source-sink tradeoffs. Although dedicated grain, forage, and silage sorghum cultivars are widely grown and bred in the industrialized world, dual use cultivars have traditionally been important in the developing world (Blümmel and Reddy 2006). Beyond grain and forage ideotypes, the specialty sorghums that receive the most investment in crop improvement currently are sweet, biomass/energy, and perennial cultivars.

Sweet sorghum has traditionally been used for fresh eating, syrup production, and recently biofuel production (Makanda et al. 2009; Murray et al. 2009). While cultivars considered sweet sorghums are always tall, have high biomass and juicy sweet stems, there are no specific scientific criteria, such as a diagnostic molecular marker or threshold sugar concentration, to formally differentiate sweet sorghums from grain sorghums. Although sweet sorghums can accumulate levels of sugar similar to sugarcane, the physiological mechanism of sugar accumulation appears to differ (Tarpley and Vietor 2007). Unlike sugarcane, maximum sucrose content occurs post-flower between the soft and hard dough stages, therefore all adapted sweet sorghums produce grain but often far less than grain sorghum (Lingle 1987; Rooney personal communication). Panicle (grain) removal from sweet sorghums does improve stem sugar concentration, but the harvestable sugar energy is not commensurate with an intact grain head (Broadhead 1973; Ritter 2007; Fortmeir and Schubert 1995). This is another line of evidence pointing to sink limitations of photosynthesis within the *Saccharinae*.

A quantitative method to examine if correlated sink tradeoffs between ideotypes are "inevitable," i.e., inseparable pleiotropic consequences of single genes, is to make a biparental cross between two specialty types and examine correlations among the progeny. If a molecular genetic map is used, correlations can be identified where QTL colocalization between traits occur. No less than six such sorghum QTL mapping populations have been investigated (Natoli et al. 2002; Bian et al. 2006; Ritter et al. 2008; Murray et al. 2008a, b; Shiringani et al. 2010; Rooney and Murray unpublished data) but only a few have reported information to formally evaluate tradeoffs. As summarized in Murray et al. (2008a), in most cases, QTL for stem sugar and grain production did not colocalize, suggesting that correlated genetic effects, if any, were below the statistical significance thresholds of these studies. The primary exception is with QTL known to affect flowering time, a trait known to confound the measurement of many other traits. As discussed further in Sect. 3.2 below, differences in the dates of flowering and maturity vary the environment the plant grows and develops in. This results in significant variation in a trait that is statistically detected as a QTL for that trait. If the parents used were truly differentiated in photosynthate allocation to different sinks, then under the classical "inevitable tradeoffs" model the high parent for a trait would likely have all the QTL alleles that increase the trait. In sorghum this prediction has usually held true at least to the statistical thresholds of QTL studies. However, an important alternative explanation is population structure.

If few direct genetic tradeoffs exist why do we see correlations between traits in ideotypes? One reason is that major effect flowering time and height genes differentiate these groups. Similarly, population structure appears to be another likely reason.

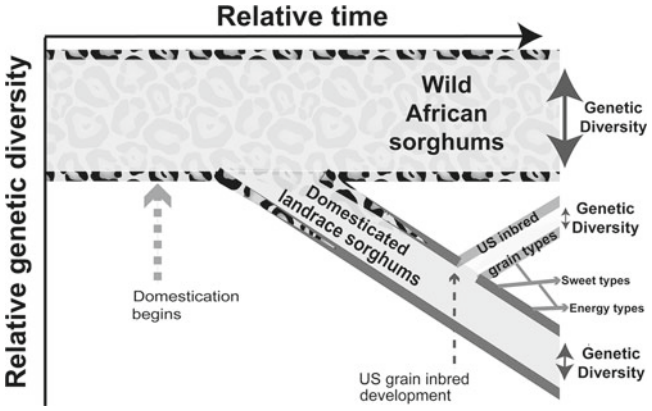

Fig. 20.1 Schematic of domestication and subsequent differentiation and substructuring of cultivars

Structure is caused when barriers to gene flow, such as geographic isolation, are combined with mutation, genetic drift and/or selection pressure. These increase the frequencies of different alleles in different subsets of the gene pool and may be independent from physiological tradeoffs. Multiple diversity panels have demonstrated differentiation of racial type, geographic origin, product type (grain, sweet, forage), and breeding program, using molecular markers (Menz et al. 2004; Menkir et al. 1997; Ritter et al. 2007; Ali et al. 2008; Casa et al. 2008; Murray et al. 2009; Shehzad et al. 2009; Wang et al. 2009). Sweet sorghum ideotypes can be discriminated with molecular markers, but this is likely caused by population structure due to breeding program and geographic origin (Ritter et al. 2007; Murray et al. 2009; Wang et al. 2009). The primary lesson from sorghum is that molecular genetic diversity can, at this point, differentiate races and geographic origins much more clearly than ideotypes. One would expect that the isolation of specialty ideotypes from other genotypes by breeding (Fig. 20.1) would eventually facilitate the accumulation of molecular signals of differentiation. However, one also would expect directional selection on genes for domestication to be easily detectable (as in maize) but it appears that these signals are not clear at that level either (Hamblin et al. 2006). Substantially more intermating between geographic and racial types followed by isolated breeding may be necessary to differentiate sorghum ideotypes based on molecular markers alone, as shown in maize by Labate et al. (1999).

When evaluating variation in new traits, working within refined versions of the sorghum diversity panels discussed above is a good initial choice. These diversity panels have now been investigated for traits such as stem sugar, flowering time, and hydrolysis potential (Vandenbrink et al. 2010). However, experience tells us that these panels should be refined to work within subpopulations for maximum statistical power whenever possible.

2.2 *Saccharum*

In contrast to most sorghum genotypes, *Saccharum* is generally a high biomass perennial that is propagated clonally, with few elite cultivars. Seed production in *Saccharum* is erratic and the seed is often very small. Accordingly, breeding and selection focus on fewer *de novo* crosses than in sorghum and maize. One would therefore expect less migration and recombination and hence more linkage disequilibrium and population substructuring. No domesticated *Saccharum* grain cultivars are known to exist and grain represents a small sink for total plant energy. Differentiation in sugarcane genotypes has primarily been towards a subset of "energy cane" types away from the "sugarcane" types as discussed in Chap. 3 by Moore et al. Energy cane types are not truly genetically differentiated as a population, often being progeny of sugarcane parent crosses that have very high biomass and low sugar content, explicable by a few major genes for stem sugar accumulation and biomass. Although no published study has looked at tradeoffs between all plant products, there have been a few biparental studies that examined aspects of both sugar content and biomass yield (never grain yield). Initial genetic study of two biparental populations by Kang et al. (1983) found that most aboveground energy sink products were either positively correlated or independent and had high heritability suggesting they could be simultaneously improved without tradeoffs. The one exception was negative correlations of stalk diameter with both brix and stalk number.

At least partially because of the genome complexity of *Saccharum*, only a few biparental QTL linkage mapping studies have been published to date. In one of the first sugarcane QTL mapping experiments, Ming et al. (2001) identified up to 36 sugar QTLs across two biparental populations. Relevant to differentiation, three of the alleles across the 36 QTLs that improved sugar came from the low sugar parent. The most closely related structural biomass traits, stalk number and stalk weight, each had positive correlation with sugar content but QTL colocalization gave mixed signals (Ming et al. 2002). Two independent investigations by Hoarau et al. (2002) and Reffay et al. (2005) also found mostly positive correlations between sugar content and measures of stalk biomass, again suggesting few tradeoffs and little differentiation between *Saccharum* ideotypes. Because no thorough studies of *Saccharum* plant products have been reported, it is difficult to know if and what additional genetic tradeoffs might exist in yield or composition of other products. Variable genetic dosage effects and population differentiation may make this even more difficult to ascertain. Unfortunately, there has been very limited analysis of *Saccharum* genetic diversity compared to the many studies in sorghum. Generally, authors have found that *Saccharum* genetic structure correlates to genome size and species but not to geographic origin (Besse et al. 1998; Pan et al. 2004; Schenck et al. 2004; Arro et al. 2006). Pan (2007) found some additional correlation to breeding program. Conversely Creste et al. (2010) did not find correlation within the various Brazilian breeding programs and Raboin et al. (2008) did not find breeding program correlations or population structure among 72 modern sugarcane clones. The failure to find differentiation in these studies is not too surprising given the

relatively short and limited history of breeding and recombination of *Saccharum*. Additionally, these results should be interpreted with caution as they are all limited by small sample sizes, use of dominant DNA markers (with few codominant markers used), and a lack of linkage data.

Thus far, one of the best demonstrations of plant sink limitations in sugarcane was conducted by Wu and Birch (2007) who, by transforming a second sucrose isomer into sugarcane, effectively doubled the sugar content of the crop. This is good evidence to suggest that within their study environment and cultivar, carbohydrate sinks are the limitation to increasing photosynthesis. As ongoing breeding efforts advance, with increasing support from DNA markers, frequencies of sugar accumulation alleles should become enriched and may begin to differentiate sugar from "energy" ideotypes.

2.3 Maize

Maize, a C4 grass closely related to the *Saccharinae*, has received heavy investments in breeding and genetics research and has a number of ideotypes and plant products that might be used as a model. Grain is by far the most important sink of the maize plant and has received most of the attention of breeding and selection programs in the developed world. In some regions there are additional long-standing interests in the genetic improvement of maize for stover, silage, and more recently biomass for cellulosic biofuel production (Carpita and McCann 2008; Lorenz et al. 2009b; Lorenzana et al. 2009). Lorenz et al. (2009a) determined that increases in yield of temperate elite maize have increased stover yield while maintaining constant harvest index (ratio of grain to aboveground biomass). Conversely, increased grain yield in high biomass tropical maize (which has undergone much less intensive breeding) has increased harvest index and decreased biomass. This suggests that harvest index may be easier to breed for first, before trying to increase grain yield and biomass simultaneously. From these results Lorenz et al. (2009a, b) concluded that stover yield can be increased independent of grain yield, with no major tradeoffs expected. Conversely, future major increases in elite temperate maize grain yield will require a commensurate increase in structural biomass.

To my knowledge, no published study has examined genetic structure between silage and grain maize or investigated QTL in a cross between ideotypes. Because grain is a critical component to both silage yield and composition, and based on the fact that few maize silage breeding programs exist, it seems unlikely that any substructuring or genetic differentiation of germplasm has yet occurred. One of the few (only?) models of true genetic differentiation due to human-based selection (as opposed to geographic isolation) is maize heterotic groups (Liu et al. 2003; Nelson et al. 2008). Impressively, substructuring for heterotic groups within individual populations of maize can be seen with only a few generations of recurrent selection; examples include the Iowa Stiff Stalk Synthetic, the Iowa Corn Borer Synthetic (Labate et al. 1999), and CIMMYT's Pool 30 (Wisser et al. 2008).

Although not widely recognized, maize also has the ability to accumulate high levels of sugar in the stem, similar to sweet sorghum and sugarcane (Stewart 1878; Setter and Meller 1984; Murray unpublished data). Archeological evidence, albeit controversial, has suggested that maize may have been domesticated for stem sugar accumulation in association with or instead of grain (Smalley and Blake 2003; Piperno et al. 2009). No attempts to breed maize for higher stem sugar have been reported, therefore genetics of stem sugar differentiation remains unknown, and appropriate material to screen has not been located.

Maize enjoys immense genetic and genomic resources due to its importance both as a grain crop and as a genetic system. Many genes that will be important to investigate have been and will continue to be first identified in maize. For instance recessive mutants of a gene called *Tie-dyed2*, which allows additional carbohydrates to accumulate in the leaves of maize plants and increase cellulose concentration, were recently identified and could be investigated for a possible role in differentiation of silage types and an additional avenue of sink storage (Baker and Braun 2008). Many of the genes for cell wall composition and architecture also have and will continue to be first identified in maize (Carpita and McCann 2008). Leveraging these as candidate genes in *Saccharinae* crops will allow faster identification of QTL conditioning important traits.

2.4 Other Grasses

Pearl millet (*Pennisetum glaucum*), a close relative of the *Saccharinae,* is fully domesticated and has been modestly researched. Pearl millet is grown primarily by subsistence farmers in Africa and Asia. Like landrace sorghums in the developing world, these farmers require both the grain for human food and the crop residue for silage to feed their animals (Yadav et al. 2003; Blümmel and Reddy 2006). Although some classical breeding studies have shown heterosis and a lack of tradeoffs between grain and biomass (fodder production), few genetic analyses of these traits have been conducted (Blümmel et al. 2003). In one reported biparental mapping population, grain yield and stover yield were positively correlated; all three grain yield and stover yield QTL colocalized with the same directional effects without tradeoffs (Yadav et al. 2002). However, it is important to note that all but one of these QTLs corresponded to flowering time loci. Due to a lack of differentiation and little research investment it is not clear what genetic structure might be present in this important subsistence crop.

No other C4 grasses have received the attention in breeding or genomics that maize, sorghum, *Saccharum*, or pearl millet has. In a new crop such as *Miscanthus* this is due to the combination of a complex genome, undomesticated nature, and lack of historical importance. The lessons we learn in creating and measuring differentiation in maize, sorghum, *Saccharum*, and millet may translate to new crops such as *Miscanthus* and will likely accelerate progress in new crop development.

3 Investigations of Yield and Composition of Sink Products

The single most important criteria in determining the agronomic and commercial success of a C4 grass cultivar is yield. Yield is best defined by the harvest mass (weight) of the product of interest per unit area of production. Thus, grain cultivars would be expected to be selected based on grain weight and dedicated cellulosic cultivars would be selected based on harvestable dry cellulosic biomass weight. Crops are commodities so there is generally some tolerance of compositional variations, treating products as if they were functionally equivalent so they can be sold simply based on weight. However, the amount of energy put into a sink by the plant and extractable by humans can differ markedly between cultivars. A clear example is the concentration of stem sugar. Sweet sorghum versus forage sorghum or sugarcane versus energy cane can have similar stalk and juice yields but differ markedly in sugar concentration and therefore for sugar yield. In grain, composition is generally less variable but can still differ substantially. There are rarely any financial incentives for improved composition such as higher protein or lipid percentage but these parameters are important in specialty cultivars and are related to the amount of energy a plant can capture and store. Assigning value to composition and quality can be time- and cost-prohibitive. In Sect. 3.3 I will cover methods for integrating and improving compositional detection and analysis.

3.1 Grain

Grain, in most agronomic crops, is the most energy-dense plant product and consequently it is the most valuable and easiest to store, move, and trade (USDOE 2007). A majority of important domesticated crops have been selected for grain as the primary plant product though it was unlikely to be the primary energy sink of their wild relatives. Grain is typically evaluated solely on yield regardless of composition. Although average composition of grain sorghum and pearl millet appear to be similar, both have higher oil/lipids and protein than maize (Burton et al. 1972). In studies screening diverse material where grain differentiated cultivars have been further investigated, a range of compositions have been identified. For instance, maize grain lipid content in one study ranged from 0.3 to over 15% (USDA Germplasm Resources Information Network database http://www.ars-grin.gov/). This variation is much greater compared to two studies of diverse sorghum accessions, where lipid content ranged from 2.4 to 7.1% of dry grain weight (de Alencar Figueiredo et al. 2006; Hooks et al. 2006). Furthermore, in commercial hybrids the starch component of maize is increasing much faster than the lipid or protein component (Scott et al. 2006; Duvick 1997). Generally, a higher percentage of lipid leads to a lower percentage of starch but lipids have approximately 2.25 times the amount of energy as starch thus increasing energy density of the grain. Reduced grain lipid percentage generally renders grain less energy dense and each individual kernel, keeping size constant, is a weaker photosynthesis sink.

A good comparison in composition between differentiated grain and non-grain types can be seen in grain sorghum versus sweet sorghum. Grain sorghums have been selected to have large seeds with acceptable nutritional quality. Yet, because sweet, forage and energy sorghum types have not been selected for food grain production, they often have a testa and spreader suggesting that tannins are present (Dykes and Rooney 2006). This is important because although condensed tannins have not been found to exceed 6% of the seed by weight they can reduce protein and starch digestibility by 10% (Leeson and Summers 1997) and decrease the overall efficiency of ethanol conversion (Wu et al. 2007). This is one example in which secondary metabolites confound the genetic variation for, in this case, starch digestibility and measures of energy density. Comparatively in maize, which has a larger share of the bioethanol market, hybrids have been specifically developed with improved digestibility such that the grain yields 2–5% more ethanol than standard commodity maize but the underlying compositional cause of this phenotype has not been reported (Nichols and Bothast 2008).

3.2 Stem Sugar and Nonstructural Biomass

Stem sugar yield per unit area is a function of two measured components: juice yield per unit area and sugar concentration in the juice. In both sugarcane and sorghum the maximum observed stem sugar concentration ranges between 20 and 25%. Although there are dry stalk grain sorghums, reported to be controlled by a single gene (Rangaswami Ayyangar 1935), there does not appear to be much quantitative variation for sorghum stalk juice content. Therefore, primary effort in increasing sorghum sugar yield will need to come from increasing the juice yield per hectare. This can be done both by increasing stalk size (diameter and/or height), and by increasing the number of stalks per hectare; both have been found to be independent or positively correlated with stem sugar concentration (Hoarau et al. 2002; Reffay et al. 2005). Multiple studies have found that major sorghum stem sugar concentration QTLs are either pleiotropic or tightly linked or epistatic with height and flowering time QTL (Ritter et al. 2008; Murray et al. 2008a; Shiringani et al. 2010).

In both sweet sorghums and sugarcane the stem sugar yield is extremely dependent on environment and maturity at harvest. With drought in sorghum, the concentration of sugar tends to increase while the juiciness tends to decrease, retaining approximately the same sugar yield. In many crops, stem sugar has been linked to stress resistance and suggests an evolutionary origin of this storage carbohydrate (Blum 1996, 1998). Once stem sugar genes are cloned it may be possible to identify the origins and mechanism of sugar accumulation through diversity and expression analyses. Though some important enzymes and genes in the pathway have been identified in sugarcane, these do not appear to have corresponding importance in sweet sorghum, potentially because of the different physiological mechanisms noted above (Lingle 1987, 1999; Tarpley and Vietor 2007; McCormick et al. 2008).

3.3 Structural Biomass

Structural biomass is of importance in both forage and cellulosic bioenergy production. Structural biomass can derive from either leaves or stems and in turn can arise from a main stalk or tillers. In sorghum, structural biomass yields can be increased by altering flowering time, height, stem diameter, ratooning, tillering, and types of staygreen and perennialism (Murray et al. 2008b, Salas Fernandez et al. 2009). Later flowering material often increases structural biomass with a potential decrease or absence of grain (Rooney et al. 2007; Murray et al. 2008b). Lack of grain filling allows the crop to go dormant during the higher drought and heat stress periods of summer. To breed specifically for this, an efficient epistatic genetic system has been identified in sorghum for seed production from biomass genotypes. This system allows inbred parents to be virtually unaffected by photoperiod so that crosses can be made in temperate summer environments to generate photoperiod sensitive progeny that generally do not flower in the temperate environments in which they were produced (Rooney 1999; Mullet et al. 2010).

Structural biomass yield and resistance to lodging are of prime importance for both forage and bioenergy goals. However, the composition and hence the leaf-to-stem ratio will differ between the two; leaves typically have more protein and cellulose, while stems have more hemicellulose and lignin. Furthermore, bioenergy cultivars may be harvested at any time whereas forage cultivars must be harvested green to maintain palatability. In terms of composition, forages need to have high cellulose and low lignin for ruminant digestibility. The same thing had additionally been thought of bioenergy cultivars but improved methods are being developed to digest lignin and hemicellulose (Holtzapple and Granda 2009). Another important difference is that in forage crops, high protein and mineral content is necessary for animal feed but undesirable in a biofuel feedstock (Casler and Vogel 1999; Jenkins et al. 1998; Wu et al. 2007). This is because protein is extremely important for animal gain but reduces cellulose digestibility and fermentation efficiency during ethanol production and creates air pollution in systems that use direct combustion. Minerals, although important to animal nutrition, can foul processing equipment (Jenkins et al. 1998).

As discussed by Vermerris et al. (Chap. 17 this book), BMR sorghum and maize lines have low lignin structural composition useful for forage and biofuel uses but have a known pleiotropic effect of decreasing biomass yield, grain yield, and increasing lodging (Oliver et al. 2005). Alternatively, compositional improvement can be made by screening diversity panels for natural variation in these traits. Variation in lignin composition found in the sweet sorghum diversity panel exceeds that reported for brown midrib mutants (Murray, unpublished data). BMR lines remain attractive to producers because have a simple identifying visual marker to select on and preserve identity. It seems likely that the three main BMR mutant loci (Pedersen et al. 2008) could be used as DNA markers to differentiate the BMR forages as a subpopulation. There are also only a few extreme photoperiod sensitivity

loci (ma5 and ma6) that will be used to differentiate the improved energy cultivars. Otherwise, differentiation of compositional ideotypes is likely to be extremely difficult with molecular markers at the present time.

3.4 Rhizomes

Close wild relatives of most of the *Saccharinae* produce rhizomes for perennial overwintering, clonal propagation, stress protection, and carbohydrate storage (Holm et al. 1977). Very little research into sink partitioning for rhizome yield, composition, or genetics has occurred in the grasses although current and/or economic impact of perennial crops and weeds is immense (Paterson 2009; Jessup and Paterson are Chap. 21). Collecting data on underground plant parts can also be extremely time-consuming and expensive. Weedy types such as Johnsongrass are the most important and best-researched *Saccharinae* rhizomatous species thus far. In one of the few studies measuring the yield of rhizomes, McWhorter (1961) found the rhizome biomass to be more than double that of the rest of the plant! Interestingly, like sugar concentration in sweet sorghum, Johnsongrass rhizomes are at a maximum soon after seed-head production (Rapp 1947; McWhorter 1961). Since they would be expected to be competing for the photosynthates at the same time, this suggests that there are few to no tradeoffs between seed and rhizome sinks.

Although studies have been few, perennial grain sorghum cultivars derived from crosses between *Sorghum bicolor* and *Sorghum halepense* have been shown to emerge 4 weeks earlier than annual sorghum (DeHaan et al. 2005), and observations of agronomic systems in Texas support that weedy *S. halepense* emerges more than 4 weeks earlier than annual sorghum. Furthermore, regrowth from established plants can set seed within a month after a hard winter kill (Murray personal observation). Because these plants are actively growing and photosynthesizing for a month before and many months after annual domesticated sorghums, they would be expected to capture significantly more solar energy. Much of this growth period would be during less stressful, cooler, and wetter periods. In the case of Johnsongrass in colder climates, rhizomes may not escape freezing and short seasons may provide less of an advantage for perennial late photosynthesis thus rhizomes may not always be advantageous and can be lost (Warwick et al. 1984; Murray unpublished data). Greater quantitative allocation to rhizomes can be seen throughout Johnsongrass worldwide. Wild collections of *S. halepense* show it to be very diverse with various investments in plant sinks and high levels of molecular variation (Burt 1974; Warwick et al. 1986; Morrell et al. 2005; Kellogg personal communication; Murray unpublished data). There is some evidence that this may be partially due to outcrossing with domesticated sorghum (Arriola and Ellstrand 1996; Morrell et al. 2005). Interestingly maize and pearl millet also have perennial wild relatives that are not invasive (Cox et al. 2002). These could be used to add an additional sink and breed a perennial maize (Shaver 1967; Westerbergh and Doebley 2004) or pearl millet, and although some investigators have tried, much more work is necessary.

4 Appropriately Evaluating Tradeoffs

When attempting to evaluate yield, composition or energy potential of a crop, genotype, or plant sink, there are multiple confounding factors that limit the ability to estimate true genetic effects. This, in turn, limits conclusions and the ability to select elite differentiated genotypes. Furthermore, to appropriately evaluate tradeoffs in these various genotypes, any and all confounding factors must be identified and eliminated.

4.1 Physiological Experiments and Modeling Plant Tradeoffs

Classical physiological studies have been conducted in numerous species using a multitude of methods to look at resource allocation and tradeoffs in plant products and life history traits (as reviewed in Obeso 2002). The majority of these studies have been conducted in wild species in which it is difficult or inappropriate to perform controlled crosses for improvement. In domesticated *Saccharinae* species, one example of this approach is measuring increases in biomass or stem sugar from the removal of panicles or tillers (Broadhead 1973; Rajendran et al. 2000; Balole 2001; Ritter 2007). While informative and tractable from the physiological standpoint, by disrupting homeostasis these methods fail to take sink-limited production into account, to get at questions of genetic differentiation of cultivars, or to guide us in how to conduct future crop improvement (Dingkuhn et al. 2007). One method to help answer these questions is the so called "path analysis" as used to quantitatively evaluate variation across traits in segregating sugarcane populations by Kang et al. (1983). Path analysis is a traditional statistical approach of identifying cause and effect relationships in breeding populations. An attractive recent approach improving on path analysis, known as ecophysiological growth models, looks at segregating populations and takes genetic by environment (G×E) interaction into account (Dingkuhn et al. 2005). These provide a better understanding of the genomics of differentiation and suggest "limiting factors" for subsequent breeding improvement. Although these growth models still remain unproven as a way to make progress, they theoretically will assist with the transition from descriptive to predictive breeding of differentiation or maximum usable energy accumulation.

4.2 Genetic Confounding Factors

Genes controlling traits that are not of primary interest can confound our ability to detect genetic potential of a seed, sugar, biomass, or rhizome-producing ideotype. These are often assumed to be due to pleiotropic effects and can be caused by genetic and/or biological epistasis or (G×E) interaction (Moore 2005). Such confounding may occur when phenotypic evaluation is conducted in different environments

although at the same physiological stage. For example, if the physiological maturity window of a population stretches over 30 days, the temperature and soil moisture status experienced by a genotype maturing on day 1 and another genotype maturing on day 30 can vary greatly. A QTL for a trait of interest might be identified that colocalizes to maturity simply because the maturity QTL caused the trait of interest to be measured in a different environment. Environmental changes caused by genetic confounding factors can be external (i.e., air temperature) or internal (i.e., cell size, cell sugar content). Among confounding environmental factors, flowering-time and height are two with the most important interaction. Flowering time, which can often be due to photoperiod sensitivity, appears to confound analysis of nearly every other trait when measured in sorghum (Murray et al. 2008b), maize, and millet (Yadav et al. 2002). Brown and Paterson (Chap. 14 this book) review classical and modern genetics of flowering time genes and interactions.

Plant stature (height) also appears to confound a variety of traits; this is in part but not solely due to the strong positive correlation of height with flowering time. Height, as discussed previously, is key to differentiated types of plant sinks in sorghum with dedicated grain types being shorter than sweet types that are in turn shorter than energy types. Ideotype height is somewhat correlated to root and stalk lodging potential which involves stalk composition and panicle size factors as well. Because of confounding by height and flowering time, the sorghum conversion program (Klein et al. 2008) and the genetic enhancement of maize program (GEM—http://www.public.iastate.edu/~usda-gem/) were established to convert tropical tall photoperiod sensitive types to dwarf nonphotoperiod sensitive types that could be more appropriately used and evaluated for grain production.

During statistical analysis, large confounding factors mask smaller sources of variation, reducing statistical power to detect the variation most of interest. Ecophysiological models attempt to account for trait effects due to factors such as flowering time and height (Letort et al. 2007) but the lack of cross-environment observations (degrees of freedom) cannot be adjusted statistically. Another approach to minimize the impact of height and flowering time is to select from a larger population only those lines that appear uniform for further screening (see Sect. 3.2).

Because height, flowering time, and other potential confounding factors are affected by many genes, it may be helpful to think of these as genetic background effects rather than the extension of simple two gene epistasis. Genetic background and epistasis are known to confound the dissection and utilization of many traits through context dependency (Podlich et al. 2004; Cooper et al. 2009). For example, across a very large nested association mapping (NAM) QTL experiment in maize, 39 QTLs were detected with the largest effect averaging less than 1.7 days (Buckler et al. 2009). In the genetic background and environments that Salvi (2007) cloned the maize flowering time gene *Vgt1*, it was responsible for 1–4 weeks of flowering time differences. Across the genetic backgrounds used in Buckler et al. (2009) it only had an average effect of a single day across environments. In sorghum, a few genes of major effect appear to control flowering time/photoperiod sensitivity, as opposed to many genes of small effect in maize, but little effort has been made to test these across different genetic backgrounds (Brown and Paterson, Chap. 13 of

this book). If we believe that genetic background effects are controlled by many small quantitative genes rather than a few major ones, the largest potential genetic confounding factors will simply be observed as population structure and could serve to differentiate ideotypes.

4.3 Appropriate Environments

Mainly due to genotype × environment interaction, in all breeding, genetics, and agronomy research, it is important to evaluate cultivars in the target environment(s) (Atlin et al. 2001). Multiple studies have shown that for many traits (in the populations they were evaluated in), variation due to environment and genetic × environmental interaction can be greater than variation due to genetics alone. The deleterious effects that unadapted material usually experiences may be due to abiotic and biotic factors such as photoperiod responsive flowering time, temperature tolerance, disease, or pests.

An extreme example is growing photoperiod sensitive sorghum in New York, where the photoperiod does not permit flowering until early autumn. Lines that do not flower will have zero grain yield and lines that begin setting seed shortly before frost will appear to have very low yield and altered composition (Murray et al. 2009). Including grain types in the same experiment, which are usually day-neutral and flower much earlier than sweet sorghum or energy types, will be confounded by the combination of maturity factors and a seasonally changing environment. If the goal is to maximize energy production of the crop it is difficult to appropriately evaluate these since the best seed-producing types are targeted for different environments than the best biomass-producing types. Evaluating both in a common garden experiment may yield erroneous results due to varying degrees of adaptation; however, there are a few methods for dealing with this (Table 20.1).

4.4 Molecular Genetics and Breeding

Molecular genetics tools have many uses in classifying and breeding grain, sugar, biomass, perennial and multi-purpose genotypes. The majority of phenotypic traits that differentiate these ideotypes are quantitative, defined by many genes with small individual effects and environmental interactions. Many of these genes have not been identified but we would expect that divergent selection on each for different ideotypes would lead to population subdivision over the long term. From classical quantitative genetics we know that complex traits involving many genes are slower to respond to selection, and accordingly will take longer to genetically differentiate ideotypes into new subpopulations than would more simply-inherited traits (i.e., BMR). From the molecular breeding standpoint it will also be more difficult to stack traits, given that each locus is likely linked to some others in trans configuration and

Table 20.1 Methods for dealing with confounded (out-of-environment) factors, such as photoperiod sensitivity in a segregating population

	Advantages	Disadvantages
1. Ignore the confounding factor	Simplest method, no additional work needed	Huge loss in power, likely to draw erroneous conclusions about underlying mechanisms or value of certain traits
2. Measure the confounding factor, harvest all samples at the same time, and use statistical analysis to partition confounding variation	Easy postharvest correction, plant and harvest all samples at the same time, consistent with traditional field management	Significant loss in statistical power. Cannot truly separate effect but can determine how much error it is causing
3. Stagger harvest based on factor (i.e., maturity/days after anthesis)	Evaluate all plants at the same maturity and same planting environment	More complex when harvesting large numbers of samples, environments at harvest differ, can alter harvest only for one confounding factor (i.e., not maturity and disease simultaneously)
4. Conduct evaluation in an environment where the confounding factor can be eliminated, i.e., tropical winter planting which will nearly eliminate photoperiod effects (Castillo-Gonzalez and Goodman 1989), or an environment that does not have disease or insect pressure	Eliminates one or a few specific confounding factors	More expensive since have to travel to additional locations, likely not relevant to growers field conditions, not all confounding factors can be eliminated and new ones may be created

linkages will need to be broken. Molecular markers in appropriate populations will assist in separating phenotypic correlations from tight linkage or pleiotropy (with traits such as flowering time) and help to differentiate new populations. Mapped QTL can be improved via marker assisted selection or, without mapping, genomic selection approaches can be used (Bernardo 2009; Heffner et al. 2009). These approaches allow QTL that improve target components (and are non-plieotropic) to be pyramided in breeding populations. An important question to be addressed is whether genomic selection is redundant to ecophysiological modeling approaches or if both can be combined in a single model to further improve genetic gain?

Finally, QTL mapping is sometimes the most promising first step to gene cloning. Once a gene has been cloned many other crop improvement approaches can be pursued. First, superior natural/native alleles for the trait can be identified and enriched in a population (Yan et al. 2010). Second, the genetic pathway for improvement will be revealed, and other gene candidates may be identified—this is especially important in cases such as stem sugar accumulation in which the pathway is

unknown. Third, gene identification will allow transformation of novel alleles into other species or individuals. Transformation technologies are initially most useful for hypothesis testing gene functions and incorporating gene diversity from other species. For example, genes for stem sugar or photoperiod sensitivity identified in sorghum might be useful to improve these traits in maize or sugarcane. Conversely, *Tie-dyed2* recessive mutants may allow additional carbohydrates to accumulate in the leaves of sorghum or *Saccharum* (Baker and Braun 2008). Finally, the cloned gene may allow the creation of novel variation through knocking out/turning off, or over-expression.

While knockout mutants in crops make great tools for genetic dissection, they often have deleterious pleiotropic effects of little use in breeding. As described by other authors in this book, a sorghum TILLING population can be of use in gene identification via loss-of-function mutants (Xin et al. 2008). However, loss-of-function mutants have rarely been useful in crop improvement. For example, as mentioned previously BMR decreases lignin but has known pleiotropic effects of decreasing biomass yield, grain yield, and increasing lodging (Oliver et al. 2005).

4.5 Near Infrared Spectroscopy

While height, flowering time and yield are straightforward to measure, one of the largest challenges of assigning value to composition and energy content is that it can be very time and cost prohibitive to do so and limited sampling can lead to a lack of power and erroneous conclusions. Although variation in product yield appears to be much greater than composition, composition differences have led to a 23% difference in whole plant theoretical ethanol yield (Murray et al. 2008b) and a 2–8% difference in measured grain digestibility and ethanol yield (Leeson and Summers 1997; Bothast and Schlicher 2005). It seems appropriate and likely that at some time in the future, cultivars will be differentiated and sold based on their composition in addition to their yield as some forages are now. Composition can especially be very costly and time-consuming to measure in large numbers such as for genetic mapping or in bulk production out of producers fields. Near infrared spectroscopy (NIRS) is one method that can, in some cases, make goals of composition analysis attainable. In brief, chemical bonds in constituents of interest absorb light at particular wavelengths in the near infrared spectrum. By calibrating NIR absorbance against known wet-chemistry values of appropriate samples, the composition of unknown samples can be predicted. Once calibrations have been developed the only cost of analysis is a few minutes of sample scanning and all calibrated components can be predicted simultaneously and instantaneously. For sorghum grain, NIRS calibrations have been successfully developed for starch, oil, protein, endosperm texture, hardness, fiber, and phosphorus (de Alencar Figueiredo et al. 2006; Hooks et al. 2006; Murray et al. 2008a, b). For leaf and stem biomass, calibrations have been developed for cellulose, hemicellulose, lignin, crude protein and starch using forage methods in sorghum and maize (Frey et al. 2004; Murray et al. 2008b) For

stem sugar, traditional brix/pol measurements used by the sugarcane and sweet sorghum industries are already very fast and fairly accurate in quantifying sucrose. However, NIRS calibrations for brix, pol, reducing sugar, component sugars (sucrose, glucose, fructose), and extractable starches have been successfully developed and look promising to improve throughput and accuracy (Valderrama et al. 2007; Rambla et al. 1997; Rooney personal communication). Rhizome and root calibrations have yet to be developed but would be extremely helpful to thoroughly quantify all plant sink products. NIRS has also been used in mutant screens to detect significant spectral deviation in composition that might be the result of novel functional polymorphisms (Vermerris et al. 2007). While genomics technology continues to decrease the cost of screening germplasm, there has been almost no decrease in phenotyping cost. Near infrared spectroscopy may help to equalize this for composition traits.

5 Advancing Crop Improvement: Further Differentiation Versus Maximum Energy Production

To successfully advance crop improvement in the *Saccharinae*, goals must be adequately defined. Traditionally, improvement has been towards one or a few products (differentiation) with yield of these products within one or a few target environments as the only goal. This has led to very productive cultivars for individual products, sugar, grain, biomass, rhizomes, etc., with narrow adaptation. These individual cultivars have often been isolated within and between breeding programs and created from a subset of geographic landraces that are already adapted. With demand for cellulosic biofuel production, carbon sequestration, and ecosystem services combined with availability of improved technology and a global breeding focus there are increasing opportunities for multiple product production that could fit across many environments. A common concern is that genetic gain in breeding for one quantitatively controlled product is already low and breeding for more than one product could result in even less gain. Instead, I suggest that by screening large and diverse germplasm panels at the beginning and finding lines that are already superior in multiple traits related to energy production, favorable alleles, and epistatic combinations might be identified. Screening must be undertaken in multiple environments and with the considerations provided in Table 20.1. Superior lines can then be crossed and developed into new populations that excel in multiple product production.

Many important biological questions remain about differentiating ideotypes versus stacking traits. Have differentiated subpopulations been maximized for traits of interest or, for example, can alleles in grain sorghum increase stem sugar in sweet sorghum? We would expect that these QTL might have smaller effects masked by larger QTL and confounding factors such as flowering time. However, many studies, beginning with Tanksley et al. (1996) and continuing as cited here, have identified favorable QTL from the unfavorable parent. If this is universally true than genetic isolation

and differentiation of subpopulations makes little sense. If not, then can we maximize harvestable energy content by stacking alleles from differentiated populations into a single cultivar? This will be possible only if there are no pleiotropic effects (i.e., flowering time), little linkage, and all products can be economically used.

6 Conclusions

Differentiation of seed, sugar, and biomass-producing genotypes in sorghum has become the norm in modern *Saccharinae* breeding and will likely occur in other species as well. Because there are few breeding programs that intermate diverse unadapted material and a lack of a conversion program such as GEM or the sorghum conversion program for grain, it seems likely that most selections will result from narrow crosses within adapted geographic subgroups. Thus, with continued low investment in breeding and genetics in *Saccharinae* crops and the complex quantitative nature of many traits of interest, it is unlikely that we will observe any type of molecular signature able to be used in differentiating the specialty cultivars in the near future.

Acknowledgments Thanks to Adam Mahan and William Rooney for reviewing this chapter in an earlier form. Also thanks to Andrew Paterson for inclusion and thoughtful editing of this chapter.

References

Ali ML, Rajewski JF, Baenziger PS, Gill KS, Eskridge KM, Dweikat I (2008) Assessment of genetic diversity and relationship among a collection of US sweet sorghum germplasm. Mol Breed 21:497–509

Anderson LE, Appleby AP, Weseloh JW (1960) Characteristics of Johnsongrass rhizomes. Weeds 8:402–406

Arriola PE, Ellstrand NC (1996) Crop-to-weed gene flow in the genus Sorghum (Poaceae): spontaneous interspecific hybridization between Johnsongrass, *Sorghum halepense*, and crop sorghum *S. bicolor*. Am J Bot 83:1153–1160

Arro JA, Veremis JC, Kimbeng CA, Botanga C (2006) Genetic diversity and relationships revealed by AFLP among a collection of Saccharum spontaneum and related species and genera. J Am Soc Sugar Cane Tech 26:101–115

Atlin GN, Cooper M, Bjørnstad Å (2001) A comparison of formal and participatory breeding approaches using selection theory. Euphytica 122:463–475

Baker RF, Braun DM (2008) Tie-dyed2 functions with tie-dyed1 to promote carbohydrate export from maize leaves. Plant Physiol 146:1085–1097

Balole TV (2001) Strategies to improve yield and quality of sweet sorghum as a cash crop for small scale farmers in Botswana. Ph.D. dissertation, University of Pretoria, Pretoria, South Africa

Bernardo R (2009) Genomewide selection for rapid introgression of exotic germplasm in maize. Crop Sci 49:419–425

Besse P, Taylor G, Carroll B, Berding N, Burner D, McIntyre CL (1998) Assessing genetic diversity in a sugarcane germplasm collection using an automated AFLP analysis. Genetica 104:143–153

Bian Y-l, Seiji Y, Maiko I, Cai HW (2006) QTLs for sugar content of stalk in sweet sorghum (Sorghum bicolor L. Moench). Agricultural Sciences in China 5:736–744
Blum A (1996) Constitutive traits affecting plant performance under stress. In: Edmeades GO, Bänziger M, Mickelson HR, Peña-Valdivia CB (eds) Developing drought and low-N tolerant maize. CIMMYT, El Batan, Mexico, pp 131–135
Blum A (1998) Improving wheat grain filling under stress by stem reserve mobilisation. Euphytica 100:77–83
Blümmel M, Reddy BVS (2006) Stover fodder quality traits for dual-purpose sorghum genetic improvement. SAT ejournal. 2(1) available online at ejournal.icrisat.org
Blümmel M, Zerbini E, Reddy BVS, Hash CT, Bidinger F, Khan AA (2003) Improving the production and utilization of sorghum and pearl millet as livestock feed: progress towards dual-purpose genotypes. Field Crops Res 84:143–158
Borlaug NE (2002) Feeding a world of 10 billion people: the miracle ahead. In Vitro Cell Dev Biol Plant 38:221–228
Bothast RJ, Schlicher MA (2005) Biotechnological processes for conversion of corn into ethanol. Applied Microbiology and Biotechnology 67:19–25
Broadhead DM (1973) Effects of deheading on stalk yield and juice quality of Rio sweet sorghum. Crop Sci 13:395–397
Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC, Goodman MM, Harjes C, Guill K, Kroon DE, Larsson S, Lepak NK, Li H, Mitchell SE, Pressoir G, Peiffer JA, Rosas MO, Rocheford TR, Romay MC, Romero S, Salvo S, Sanchez Villeda H, da Silva HS, Sun Q, Tian F, Upadyayula N, Ware D, Yates H, Yu J, Zhang Z, Kresovich S, McMullen MD (2009) The genetic architecture of maize flowering time. Science 325:714–718
Burt GW (1974) Adaptation of Johnson grass. Weed Sci 22:59–63
Burton GW, Wallace AT, Rachie KO (1972) Chemical composition and nutritive value of pearl millet (Pennisetum typhoides (Burm.) Stapf and E. C. Hubbard) grain. Crop Sci 12:187–188
Carpita NC, McCann MC (2008) Maize and sorghum: genetic resources for bioenergy grasses. Trends Plant Sci 13:415–420
Casa AM, Pressoir G, Brown PJ, Mitchell SE, Rooney WL, Tuinstra MR, Franks CD, Kresovich S (2008) Community resources and strategies for association mapping in sorghum. Crop Sci 48:30–40
Casler MD, Vogel KP (1999) Accomplishments and impact from breeding for increased forage nutritional value. Crop Sci 39:12–20
Castillo-Gonzalez F, Goodman MM (1989) Agronomic evaluation of Latin American maize accessions. Crop Sci 29:853–861
Cooper M, van Eeuwijk F, Hammer GL, Podlich DW, Messina C (2009) Modeling QTL for complex traits: detection and context for plant breeding. Curr Opin Plant Biol 12:231–240
Cox TS, Bender M, Picone C, Van DL, Tassel HJB, Brummer EC, Zoeller BE, Paterson AH, Jackson WW (2002) Breeding perennial grain crops. Crit Rev Plant Sci 21:51–91
Creste S, Accoroni KAG, Pinto LR, Vencosvskv R, Gimenes MA, Xavier MA et al (2010) Genetic variability among sugarcane genotypes based on polymorphism in sucrose metabolism and-drought tolerance genes. Eupytica 172:435–446
de Alencar Figueiredo LF, Davrieux F, Fliedel G, Rami JF, Chantereau J, Deu M, Courtois B, Mestres C (2006) Development of NIRS equations for food grain quality traits through exploitation of a core collection of cultivated sorghum. J Agric Food Chem 54:8501–8509
DeHaan LR, Van Tassel DL, Cox TS (2005) Perennial grain crops: a synthesis of ecology and plant breeding. Renewable Agricultural Food Systems 20:5–14
Dhugga KS (2007) Maize biomass yield and composition for biofuels. Crop Sci 47:2211–2227
Dingkuhn M, Luquet D, Quilot B, de Reffye P (2005) Environmental and genetic control of morphogenesis in crops: towards models simulating phenotypic plasticity. Aust J Agr Res 56:1289–1302
Dingkuhn M, Luquet D, Clément-Vidal A, Tambour L, Kim HK, Song YH (2007) Is plant growth driven by sink regulation? In: Spiertz JHJ, Struik PC, Van Laar HH (eds) Scale and complexity in plant systems research: geneplant-crop relations., pp 157–170

Duvick DN (1997) What is yield? In: Edmeades GO, Bänziger M, Mickelson HR, Peña-Valdivia CB (eds) Developing drought and low-N tolerant maize. CIMMYT (Centro Internacional de Mejoramiento de Maíz y Trigo [International Maize and Wheat Improvement Center], El Batan, Mexico, pp 332–335

Dykes L, Rooney LW (2006) Sorghum and millet phenols and antioxidants. J Cereal Sci 44:236–251

Fortmeir R, Schubert S (1995) Storage of non-structural carbohydrates in sweet sorghum (Sorghum bicolor L. Moench): comparison of sterile and fertile lines. J Agron Crop Sci 175:189–193

Frey TJ, Coors JG, Shaver RD, Lauer JG, Eilert DT, Flannery PJ (2004) Selection for silage quality in the Wisconsin Quality Synthetic and related maize populations. Crop Sci 44:200–1208

Givnish TJ (1982) On the adaptive significance of leaf height in forest herbs. Am Nat 120:353–381

Hamblin MT, Casa AM, Sun H, Murray SC, Paterson AH, Aquadro CF, Kresovich S (2006) Challenges of detecting directional selection after a bottleneck: lessons from *Sorghum bicolor*. Genetics 173:953–964

Harlan JR, deWet JWJ (1972) A simplified classification of sorghum. Crop Sci 12:172–176

Heffner EL, Sorrells ME, Jannink J-L (2009) Genomic selection for crop improvement. Crop Sci 49:1–12

Hoarau JY, Grivet L, Offman B, Raboin L-M, Diorflar J-P, Payet J, Hellman M, D'Hont A, Glaszmann J-C (2002) Genetic dissection of a modern sugarcane cultivar (Saccharum spp.). II. Detection of QTLs for yield components. Theor Appl Genet 105:1027–1037

Holm LG, Donald P, Pancho JV, Herberger JP (1977) The World's worst weeds: distribution and biology. The University Press of Hawaii, Honolulu, Hawaii, 609 p

Holtzapple MT, Granda CB (2009) Carboxylate platform: the MixAlco process part 1: comparison of three biomass conversion platforms. Appl Biochem Biotechnol 156:95–106

Hooks T, Pedersen JF, Marx DB, Vogel KP (2006) Variation in the U.S. photoperiod insensitive sorghum collection for chemical and nutritional traits. Crop Sci 46:751–757

Jenkins BM, Baxter LL, Miles TR Jr, Miles TR (1998) Combustion properties of biomass. Fuel Process Technol 54:17–46

Kang MS, Miller JD, Tai PYP (1983) Genetic and phenotypic path analyses and heritability in sugarcane. Crop Sci 23:643–647

Klein RR, Mullet JE, Jordan DR, Miller FR, Rooney WL, Menz MA, Franks CD, Klein PE (2008) The effect of tropical sorghum conversion and inbred development on genome diversity as revealed by high-resolution genotyping. Crop Sci 48(S1):S12–S26

Labate J, Lamkey KR, Lee M, Woodman WL (1999) Population genetics of increased hybrid performance between two maize populations under reciprocal recurrent selection. In: Coors J, Pandey S (eds) Genetics and exploitation of heterosis in crops. CIMMYT, Mexico City, pp 127–137, ASA, Madison, WI

Leeson S, Summers JD (1997) Commercial poultry nutrition, 2nd edn. Guelph, Canada, University Books

Letort V, Mahe P, Cournède P-H, de Reffye P, Courtois B (2007) Optimizing plant growth model parameters for genetic selection based on QTL mapping. In: Fourcaud T, Zhang X (eds) Plant growth modeling, simulation, visualization and their applications. IEEE Computer Society, Los Alamitos, California, pp 16–21

Lingle SE (1987) Sucrose metabolism in the primary culm of sweet sorghum during development. Crop Sci 27:1214–1219

Lingle SE (1999) Sugar metabolism during growth and development in sugarcane internodes. Crop Sci 39:480–486

Liu K, Goodman M, Muse S, Smith JS, Buckler E, Doebley J (2003) Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. Genetics 165:2117–2128

Loehle C (1987) Partitioning of reproductive effort in clonal plants: a benefit-cost model. Oikos 49:199–208

Lorenz A, Anex R, Isci A, Coors J, de Leon N, Weimer P (2009a) Forage quality and composition measurements as predictors of ethanol yield from maize (Zea mays L.) stover. Biotechnol Biofuels 2:5

Lorenz A, Coors J, de Leon N, Wolfrum E, Hames B, Sluiter A, Weimer PJ (2009b) Characterization, genetic variation, and combining ability of maize traits relevant to the production of cellulosic ethanol. Crop Sci 49:85–98

Lorenzana RE, Lewis MF, Jung K, Jung H-JG, Bernardo R (2009) Quantitative trait loci and trait correlations for maize stover cell wall composition and glucose release for cellulosic ethanol. Crop Sci 50:541–555

Makanda I, Pangirayi Tongoonaa P, Dereraa J (2009) Combining ability and heterosis of sorghum germplasm for stem sugar traits under off-season conditions in tropical lowland environments. Field Crop Res 114:272–279

Marcelis L, Heuvelink E (2007) Concepts of modelling carbon allocation among plant organs. In: Vos J, Marcelis L, de Visser P, Struijk P, Evers J (eds) Functional–structural plant modeling in crop production, Wageningen UR Frontis Series 22. Springer, Dordrecht, The Netherlands, pp 103–111

McCormick AJ, Cramer MD, Watt DA (2008) Changes in photosynthetic rates and gene expression of leaves during a source–sink perturbation in sugarcane. Ann Bot 101:89–102

McWhorter CG (1961) Carbohydrate metabolism of johnsongrass as influenced by seasonal growth and herbicide treatments. Weeds 9:563–568

Menkir A, Goldsbrough P, Ejeta G (1997) RAPD based assessment of genetic diversity in cultivated races of sorghum. Crop Sci 37:564–569

Menz M, Klein R, Unruh N, Rooney W, Klein P, Mullet J (2004) Genetic diversity of public inbreds of sorghum determined by mapped AFLP and SSR markers. Crop Sci 44:1236–1244

Ming R, Liu SC, Moore PH, Irvine JE, Paterson AH (2001) QTL analysis in a complex autopolyploid: genetic control of sugar content in sugarcane. Genome Res 11:2075–2084

Ming R, Wang YW, Draye X, Moore PH, Irvine JE, Paterson AH (2002) Molecular dissection of complex traits in autopolyploids: mapping QTLs affecting sugar yield and related traits in sugarcane. Theor Appl Genet 105:332–345

Moore JH (2005) A global view of epistasis. Nat Genet 37:13–14

Morrell PL, Williams-Coplin TD, Lattu AL, Bowers JE, Chandler JM, Patterson AH (2005) Crop-to-weed introgression has impacted allelic composition of Johnsongrass populations with and without recent exposure to cultivated sorghum. Mol Ecol 14:2143–2154

Mullet JE, Rooney WL, Klein PE, Morishige D, Murphy R, Brady JA (2010) Discovery and utilization of sorghum genes (MA5/MA6). US Patent 20100024065

Murray SC, Sharma A, Rooney WL, Klein PE, Mullet JE, Mitchell SE, Kresovich S (2008a) Genetic improvement of sorghum as a biofuel feedstock I: quantitative loci for stem sugar and grain nonstructural carbohydrates. Crop Sci 48:2165–2179

Murray SC, Rooney WL, Mitchell SE, Sharma A, Klein PE, Mullet JE, Kresovich S (2008b) Genetic improvement of sorghum as a biofuel feedstock II: quantitative loci for stem and leaf structural carbohydrates. Crop Sci 48:2180–2193

Murray SC, Rooney WL, Hamblin MT, Mitchell SE, Kresovich S (2009) Sweet sorghum diversity and association mapping for brix and height. Plant Genome 2:48–62

Natoli A, Gorni C, Chegdani F, Ajmone Marsan P, Colombi C, Lorenzoni C, Marocco A (2002) Identification of QTLs associated with sweet sorghum quality. Maydica 47:311–322

Nelson PT, Coles ND, Holland JB, Bubeck DM, Smith S, Goodman MM (2008) Molecular characterization of maize inbreds with expired U.S. Plant Variety Protection. Crop Sci 48:1673–1685

Nichols NN, Bothast RJ (2008) Production of ethanol from grain. In: Vermerris W (ed) Genetic improvement of bioenergy crops. Springer, New York, pp 75–88

Obeso JR (2002) The costs of reproduction in plants. New Phytologist 155:321–348

Oliver AL, Pedersen JF, Grant RJ, Klopfenstein TJ (2005) Comparative effects of the sorghum bmr-6 and bmr-12 genes I: forage sorghum yield and quality. Crop Sci 45:2234–2239

Pan YB (2007) Genetic diversity and phylogenetic relationships among sugarcane and related species determined from microsatellite DNA data. J Am Soc Sugar Cane Technol 27:57

Pan YB, Burner DM, Legendre BL, Grisham MP, White WH (2004) An assessment of the genetic diversity within a collection of *S. spontaneum* with RAPD-PCR. Genet Resour Crop Evol 51:895–903

Paterson AH (2009) Rhizomatousness: genes important for a weediness syndrme. In Stewart Jr CN. Weeding and Invasive. Plant Genomes Wiley, pp 99–109
Paul MJ, Foyer CH (2001) Sink regulation of photosynthesis. J Exp Bot 52:1383–1400
Pedersen JF, Toy JJ, Funnell DL, Sattler SE, Oliver AL, Grant RA (2008) Registration of BN611, AN612, BN612, and RN613 sorghum genetic stocks with stacked bmr-6 and bmr-12 genes. J Plant Registrations 2:258–262
Piperno DR, Ranere AJ, Holst I, Iriarte J, Dickau R (2009) Starch grain and phytolith evidence for early ninth millennium B.P. maize from the Central Balsas River Valley, Mexico. Proc Natl Acad Sci U S A 106:5019–5024
Podlich DW, Winkler CR, Cooper M (2004) Mapping as you go: an effective approach for marker-assisted selection of complex traits. Crop Sci 44:1560–1571
Raboin LM, Pauquet J, Butterfield M, D'Hont A, Glaszmann JC (2008) Analysis of genome-wide linkage disequilibrium in the highly polyploid sugarcane. Theor Appl Genet 116:701–714
Rajendran C, Ramamoorthy K, Backiyarani S (2000) Effect of deheading on juice quality characteristics and sugar yield of sweet sorghum. J Agron Crop Sci 185:23–26
Rambla FJ, Garrigues S, de la Guardia M (1997) PLS-NIR determination of total sugar, glucose, fructose and sucrose in aqueous solutions of fruit juices. Anal Chim Acta 334:41–53
Rangaswami Ayyangar GN (1935) Juiciness and sweetness in sorghum stalks. Madras Agric J 23: 350–352
Rapp KE (1947) Carbohydrate metabolism of Johnsongrass. Agron J 39:869–873
Reffay N, Jackson PA, Aitken KS, Hoarau JY, D'Hont A, Besse P, McIntyre CL (2005) Characterisation of genome regions incorporated from an important wild relative into Australian sugarcane. Mol Breed 15:367–381
Reynolds MP, Rajaram S, Sayre KD (1999) Physiological and genetic changes of irrigated wheat in the post-green revolution period and approaches for meeting projected global demand. Crop Sci 39:1611–1621
Reynolds MP, Pellegrineschi A, Skovmand B (2005) Sink-limitation to yield and biomass: a summary of some investigations in spring wheat. Ann Appl Biol 146:39–49
Ritter KB (2007) An investigation into the genetics and physiology of sugar accumulation in sweet sorghum as a potential model for sugarcane. Ph.D. diss., University of Queensland, School of Land, Crop and Food Sciences. St Lucia, Australia
Ritter KB, McIntyre CL, Godwin ID, Jordan DR, Chapman SC (2007) An assessment of the genetic relationship between sweet and grain sorghums, within *Sorghum bicolor* ssp. bicolor (L.) Moench, using AFLP markers. Euphytica 157:161–176
Ritter KB, Jordan DR, Chapman SC, Godwin ID, Mace ES, McIntyre CL (2008) Identification of QTL for sugar-related traits in a sweet × grain sorghum (*Sorghum bicolor* L. Moench) recombinant inbred population. Mol Breed 22:367–384
Rooney WL (1999) Genetic control of a photoperiod-sensitive response in *Sorghum bicolor* (L.) Moench. Crop Sci 39:397–400
Rooney W, Blumenthal J, Bean B, Mullet J (2007) Designing sorghum as a dedicated bioenergy feedstock. Biofuels Bioprod Bioref 1:147–157
Salas Fernandez MG, Becraft PW, Yin Y, Lubberstedt T (2009) From dwarves to giants? Plant height manipulation for biomass yield. Trends Plant Sci 14:454–461
Salvi S (2007) Conserved non-coding genomic sequences associated with a flowering-time quantitative trait locus in maize. Proc Natl Acad Sci U S A 104:11376–11381
Schenck S, Crepeau MW, Wu KK, Moore PH, Yu Q, Ming R (2004) Genetic diversity and relationships in native hawaiian *Saccharum officinarum* sugarcane. J Hered 95(4):327–331
Scott MP, Edwards JW, Bell CP, Schussler JR, Smith JS (2006) Grain composition and amino acid content in maize cultivars representing 80 years of commercial maize varieties. Maydica 51:417–423
Setter TL, Meller VH (1984) Reserve carbohydrate in maize stem: [14C]glucose and [14C]sucrose uptake characteristics. Plant Physiol 75:617–622
Shaver DL (1967) Perennial maize. J Hered 58:270–273

Shehzad T, Okuizumi H, Kawase M, Okuno K (2009) Development of SSR-based sorghum (*Sorghum bicolor* (L.) Moench) diversity research set of germplasm and its evaluation by morphological traits. Genet Resour Crop Evol 56:809–827

Shiringani AL, Frisch M, Friedt W (2010) Genetic mapping of QTLs for sugar-related traits in a RIL population of *Sorghum bicolor* L. Moench. Theor Appl Genet 121:323–336

Smalley S, Blake M (2003) Sweet beginnings: stalk sugar and the domestication of maize. Curr Anthropol 44:675–703

Stewart FL (1878) Sugar made from maize and sorghum. The Republic Company, Washington D.C., 106p

Tanksley SD, Grandillo S, Fulton TM, Zamir D, Eshed Y, Petiard V, Lopez J, Beck-Bunn T (1996) Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its relative *L. pimpinellifolium*. Theor Appl Genet 92:213–224

Tarpley L, Vietor DM (2007) Compartmentation of sucrose during radial transfer in mature sorghum culm. BMC Plant Biol 7:33

USDOE (2007) Roadmap for bioenergy and biomass products in the United States. U.S. Department of Energy Office of Energy Efficiency and Renewable Energy, Office of the Biomass Program, Washington, D.C.http://www1.eere.energy.gov/biomass/pdfs/obp_roadmapv2_web.pdf

Valderrama P, Braga JWB, Poppi RJ (2007) Validation of multivariate calibration models in the determination of sugar cane quality parameters by near infrared spectroscopy. J Braz Chem Soc 18:259

Vandenbrink JP, Delgado MP, Frederick JR, Feltus FA (2010) A sorghum diversity panel biofuel feedstock screen for genotypes with high hydrolysis yield potential. Ind Crops Prod 31:444–448

Vermerris W, Saballos A, Ejeta G, Mosier NS, Ladisch MR, Carpita NC (2007) Molecular breeding to enhance ethanol production from corn and sorghum stover. Crop Sci 47(Suppl 3):142–153

Wang ML, Zhu C, Barkley NA, Chen Z, Erpelding JE, Murray SC, Tuinstra MR, Tesso T, Pederson GA, Yu J (2009) Genetic diversity and population structure analysis of accessions in the US historic sweet sorghum collection. Theor Appl Genet 120:13–23

Warwick SI, Thompson BK, Black LD (1984) Population variation in *Sorghum halepense*, Johnson grass, at the northern limits of its range. Can J Bot 62:1781–1790

Warwick SI, Phillips D, Andrews C (1986) Rhizome depth: the critical factor in winter survival of *Sorghum halepense* (L.) Pers. (Johnson grass). Weed Res 26:381–387

Westerbergh A, Doebley J (2004) Quantitative trait loci controlling phenotypes related to the perennial versus annual habit in wild relatives of maize. Theor Appl Genet 109:1544–1553

Williams RD, Ingber BF (1977) The effect of intraspecific competition on the growth and development of Johnsongrass under greenhouse conditions. Weed Sci 25:293–297

Wisser RJ, Murray SC, Kolkman JM, Ceballos H, Nelson RJ (2008) Selection mapping of loci for quantitative disease resistance in a diverse maize population. Genetics 180:583–599

Wu L, Birch RG (2007) Doubled sugar content in sugarcane plants modified to produce a sucrose isomer. Plant Biotechnol J 5:109–117

Wu X, Zhao R, Bean SR, Seib PA, McLaren JS, Madl RL, Tuinstra M, Lenz MC, Wang D (2007) Factors impacting ethanol production from grain sorghum in the dry-grind process. Cereal Chem 84:130–136

Xin Z, Wang ML, Barkley NA, Burow G, Franks C, Pederson G, Burke J (2008) Applying genotyping (TILLING) and phenotyping analyses to elucidate gene function in a chemically induced sorghum mutant population. BMC Plant Biol 8:103

Yadav RS, Hash CT, Bidinger FR, Cavan GP, Howarth CJ (2002) Quantitative trait loci associated with traits determining grain and stover yield in pearl millet under terminal drought-stress conditions. Theor Appl Genet 104:67–83

Yadav RS, Bidinger FR, Hash CT, Yadav YP, Yadav OP, Bhatnagar SK, Howarth CJ (2003) Mapping and characterisation of QTL×E interactions for traits determining grain and stover yield in pearl millet. Theor Appl Genet 106:512–520

Yan J, Kandianis CB, Harjes CE, Bai L, Kim EH, Yang X, Skinner DJ, Fu Z, Mitchell S, Li Q, Fernandez MG, Zaharieva M, Babu R, Fu Y, Palacios N, Li J, Dellapenna D, Brutnell T, Buckler ES, Warburton ML, Rocheford T (2010) Rare genetic variation at *Zea mays* crtRB1 increases beta-carotene in maize grain. Nat Genet 42:322–327

Chapter 21
Perennialism and Weediness in the Saccharinae

Russell W. Jessup

Abstract The apparent contrast in demands for sustainability and productivity in modern agriculture may be reconcilable via the genetic difference in degree between perennialism and weediness. Effective use of genomic tools may soon allow precise metering of the genetic components required to ensure perennial life status, minimize weediness, and maximize crop yields. The Saccharinae includes both model genomic species and leading food, feed, forage, fuel, and industrial crops upon which translational technologies can be evaluated and deployed. In particular, a balance between the high agricultural productivity demanded in order to minimize land requirements and perennial growth habits necessary to ensure sustainable cropping systems is sought.

Keywords Perennialism • Weediness • *Sorghum halepense* • *Saccharum spontaneum* • *Miscanthus* • *Microstegium* • Rhizomes • Tillering • Polyploid

1 Introduction

The world's terrestrial biomes consist primarily of perennial plants in polyculture (Chiras and Reganold 2004), yet more than two-thirds of global croplands contain monocultures of annual crops. Such annual crops require intensive management, well-timed inputs, favorable weather during narrow time windows, and have shorter growing seasons and less-extensive root systems that provide diminished protection against soil erosion, water and nutrient leaching, subterranean carbon losses, and biotic/abiotic pressures (Glover 2005). Perennial plants are in contrast efficient soil,

R.W. Jessup (✉)
Department of Soil & Crop Sciences, Texas A&M University,
370 Olsen Blvd., 2474 TAMU, College Station, TX 77843-2474, USA
e-mail: rjessup@tamu.edu

A.H. Paterson (ed.), *Genomics of the Saccharinae*, Plant Genetics and Genomics: Crops and Models 11, DOI 10.1007/978-1-4419-5947-8_21,

nutrient, and water managers consisting of substantial subterranean structures that have significant yet unutilized potential for repartitioning into aboveground grain and biomass.

The North American flora consists of approximately 5,000 nonnative and 17,000 native higher plants (Morin 1995; Morse et al. 1995). Despite this influx of 23% alien species in four centuries of contact with other parts of the world, only between 67 and 104 plant taxa are responsible for 90% of the economic damage caused by weeds (Holm et al. 1997). Invasive species number approximately 550 worldwide and represent an even smaller component of the world's estimated 250,000 species of flowering plants (Weber 2003). Williamson and Fitter (1996) proposed the "Tens Rule," stating that on average one in ten introduced species will escape cultivation, one in ten of these will become naturalized as self-sustaining populations, and one in ten of these will become invasive. Weeds generally have a high level of phenotypic plasticity, perhaps derived from heterozygosity via polyploidy, that provides valuable adaptive flexibility.

Although detailed studies are lacking, there is substantial evidence that weeds can provide valuable agroecological services by increasing plant diversity (Swift and Anderson 1993), nutrient cycling and soil quality (Jordan and Vatovec 2003), symbiotic bacterial populations (Sturz et al. 2001), and arthropod pest control (Bugg 1992).

2 Perennialism

Perennialism is a function of meristem indeterminacy in concert with the processes of programmed cell death utilized by plants to generate differentiated morphologies. This balance between successive rounds of tissue growth and death is a major determinant of perenniality, and for a plant to successfully perenniate the apical meristem of at least one shoot axis must remain indeterminate beyond the first growing season. In this context, perennials can be understood in terms of the extent to which apical meristems remain ahead of the succeeding wave of cell death.

Other than bamboo, perennial grasses are polycarpic perennials and survive after the first flowering to do so annually (Thomas et al. 2000). In addition to a generally longer juvenile phase, differential competence—the ability for some apical meristems to remain vegetative while others flower—in apical meristems within an individual plant is also a distinguishing characteristic of polycarpic perennials (Thomas et al. 2000). A degree of juvenility and differential floral competence in apical meristems of a mature plant are therefore the two major prerequisites for perenniality.

Perenniality in grasses may result from vegetative organs such as root pieces, rhizomes, stolons, or axillary nodal buds. Among these structures, rhizomes are one of the key distinguishing features between cultivated cereals and their wild relatives that permit survival in harsh environments. Rhizomes are thus a means of propagation and persistence that may be either useful or undesirable depending on the circumstances.

Perennialism affords many adaptations to variable environments. Annual precipitation losses in soil profiles can be five times lower in perennials in comparison

with annual cropping systems (Randall et al. 1997), which can lose up to 45% of total water inputs (Dinnes et al. 2002). Moreover, inefficient water and nutrient use by annuals may result in water-logged soils, high fertilizer costs, decreased productivity, and soil loss. A 100-year comparison between permanent grassland and annual cropping systems (Gantzer et al. 1990) revealed 30% greater topsoil retention in the perennial system. The perennials were 54 times more effective in controlling erosion than annual crops, with cover management (perennial versus annual) 35 and 28 times more significant than soil erodibility and terrain slope in explaining the differences.

Less than 50% of applied nitrogen fertilizer is captured by annual crops (Tilman et al. 2002), whereas some perennials can be fertilized at a rate of 200 kg N/ha/year and lose less than 1% to leaching (Paustian et al. 1990). Glover et al. (2010) concluded that unfertilized perennial grasslands could provide comparable levels of harvested N in biomass to adjacent high-input wheat fields' harvested grain. Specifically, over a 75-year period approximately 26% more N per hectare was harvested from the unfertilized perennial grasslands than from annual crop fields. In addition, unfertilized perennial pastures maintained greater levels of soil carbon, reduced nitrogen leaching losses, and had substantially lower energy requirements despite the large yearly rates of nitrogen removal.

Herbicides are the most commonly used type of pesticide and account for almost half of all pesticides sold (Yudelman et al. 1998). Prairie restoration to perennial species, in contrast, has been shown to virtually eliminate herbicide use almost completely and reduce weed biomass by 94% (Blumenthal et al. 2003).

Belowground to a depth of 1 m, root biomass in perennial grasslands can approach seven times that in the annual cereal crops. In addition, microbial biomass in the top meter of surface soils are nearly three times as great in grasslands compared to cultivated fields. Abundance of nonparasitic nematode–plant associates is also significantly higher in perennial grasslands than in croplands, resulting in reduced yield losses due to plant-feeding nematodes and increased nutrient conservation due to decomposition pathway shifts favoring fungi over bacteria (Culman et al. 2010). Perennial crops can also store substantially more carbon in the soil (320–440 kg/ha/year) than annual crops (0–300 kg/ha/year) (Robertson et al. 2000). This ability for greater carbon storage and reduced inputs further substantiates the potential of perennial crops towards mitigation of climate change. Their net values for greenhouse gas emissions are negative, having been estimated at −200 to −1,050 kg of carbon dioxide (CO_2) equivalents per hectare per year, as compared with positive potentials of 410–1,140 kg/ha/year for annual crops (Robertson et al. 2000).

3 Weediness

Weed species span the entire plant kingdom, yet are significantly overrepresented in just 14 families. Despite only representing 16% of all plant families, monocot families make up eight (57%) of the overrepresented families. This is significantly more

weed-rich monocot families than expected by chance, with the Poaceae perhaps the most spectacularly overrepresented among both serious and widespread weeds (Daehler 1998). A large proportion of agricultural land worldwide, however, is used for cultivating crop species within the Poaceae (rice, wheat, corn, sugarcane, etc.). This suggests that one possible explanation for the overrepresentation of the Poaceae among weeds might simply be ecological similarity, i.e., similar growing conditions favor both Poaceae crop and weed species. Comparisons of weed flora across crops do not support this hypothesis, however, as Poaceae species have been found to be weeds in 44% of non-Poaceae crops as compared to 33% of Poaceae crops (Holm et al. 1977).

The incidence of polyploidy among the most invasive weed species is high, with at least 16 of the top 18 species in the Holm et al. (1977) list of the world's worst weeds being polyploid (Brown and Marshall 1981). This trend follows in recently formed allopolyploid hybrids found to be successful weed species (Ellstrand and Schierenbeck 2000; Soltis and Soltis 2000). Thus, invasiveness in plants might be predictable given that the natural occurrence of wide hybrids due to the breakdown of reproductive barriers due to illegitimate hybridization is concentrated within certain families and genera. For example, the Poaceae is second only to the Asteraceae in frequency of such events (Ellstrand et al. 1996). The resulting polyploid hybrids tend to have greater fitness than the diploid progenitor species, possibly due to increased heterozygosity and reduced inbreeding depression (Soltis and Soltis 2000). Additional contributions to genetic variation could arise from multiple origins of polyploidy and the prevalence of genomic rearrangements within allopolyploids. In sterile and asexual allopolyploids, evolutionary benefits of uniformity could coincide with fixed heterosis (Soltis and Soltis 2000).

Despite the above-mentioned trends, the likelihood of a given plant species becoming a noxious weed is difficult to predict based on any one trait, or small number of traits, and attributes contributing to weediness vary across taxa. As one example, seed mass is generally higher for several invasive species in their invasive rather than native ranges (Buckley et al. 2003; Daws et al. 2007). From comparisons to species with weediness potential in the Saccharinae (Table 21.1), it is clear increased that seed mass would be contrary to weediness in this group where long-distance seed dispersal mechanisms via lightweight caryopses designed for wind-borne distribution are omnipresent. With this caveat, traits with potential for increasing the weediness of Saccharinae species are given in Table 21.1. Detailed descriptions of selected, known species follow in a more defensible, case-study basis.

3.1 Sorghum halepense

Johnsongrass (*Sorghum halepense* (L.) Pers.) is a perennial, tetraploid ($2n=4x=40$) species combining the genomes of *Sorghum bicolor* and diploid, perennial *Sorghum propinquum* (Paterson et al. 1995). It is a perennial and notorious weed worldwide, including naturalization throughout much of the United States and Canada.

Table 21.1 Weediness characteristics of Saccharinae and model crop species

	Target species										Crop Species	
Weediness characteristic	*Sorghum halepense*	*Saccharum spontaneum*	*Saccharum arundinaceum*	*Saccharum officinarum*	*Saccharum ravennae*	*Miscanthus sinensis*	*Miscanthus sacchariflorus*	*Miscanthus ×giganteus*	*Miscanthus floridulus*	*Micro-stegium vimineum*	*Sorghum bicolor*	*Zea mays*
Rapid seedling growth	+	−	−	−	−	−	−	−	−	+	+	+
Short vegetative phase	−	−	−	−	−	−	−	−	−	−	−	−
Indeterminant	+	+	+	+	+	+	+	+	+	−	−	−
High seed production	+	−	−	−	−	−	−	−	−	+	−	−
Long-distance seed dispersal	+	+	+	+	+	+	+	+	+	+	−	−
Rhizomatous	+	+	−	−	−	−	+	+	−	−	−	−
Nodal axillary buds	+	+	+	+	+	−	+	+	−	+	−	−
Deep root system	+	+	+	+	+	+	+	+	+	−	−	−
Seed dormancy	+	+	+	+	+	+	+	−	+	+	−	−
Environmental plasticity	+	+	+	+	+	+	+	+	+	+	−	−

Across most of this range Johnsongrass perenniates via rhizomes capable of producing rapidly growing shoots at least 1 month earlier in the spring than emergence of annual sorghum seed. The increased range of *S. halepense* has occurred recently and rapidly in North America, including northward expansion by 5° latitude between 1926 and 1979; however, most northern populations are annual due to winterkill of rhizomes (Warwick et al. 1986). Johnsongrass rhizomes store starch, similar to *S. propinquum*, and consequently lack the cold-hardiness of fructosan storing temperate, perennial grasses (Monoghan 1979). Annual populations of *S. halepense* contrast with perennial populations in several respects, including larger seed, increased seedling emergence and growth, earlier flowering, and fewer rhizomes (Warwick et al. 1984). The required genetic variation for this large ecological shift northward invading *S. halepense* may have derived by introgression from domesticated *S. bicolor* (Warwick et al. 1984).

No historical attempts to intentionally domesticate *S. halepense* are apparent. Early research on hybrids between diploid sorghum and johnsongrass (Hadley 1953, 1958; Hadley and Mahan 1956) produced 30-chromosome hybrids that were male-sterile but could be successfully backcrossed to the sorghum parent and 40-chromosome hybrids derived from fertilization of unreduced female gametes in the sorghum parent. The 30-chromosome plants were also more strongly rhizomatous. Induced tetraploid *S. bicolor* lines increased hybridization efficiency with *S. halepense* (Casady and Anderson 1952; Piper and Kulakow 1994), and breeding efforts to develop perennial grain sorghum through *S. bicolor* × *S. halepense* hybridization have been ongoing at the Land Institute since the 1980s (Piper and Kulakow 1994). Such hybrids are synonymously referred to as Columbus grass (*Sorghum* × *almum* Parodi) (Davis and Edye 1959) and show potential for selection of less-invasive rhizomes toward developing productive but non-weedy perennial sorghums. Comparatively, several improved Columbus grass forage sorghum cultivars have been released without becoming serious weeds.

3.2 Saccharum spp.

Commercial sugarcane (*Saccharum officinarum* L.) cultivars are interspecific hybrids (see Moore et al. Chap. 3) that have lost many of the critical weediness attributes of *Saccharum spontaneum*, the other parental species (Holm et al. 1997). Most commercial cultivars of sugarcane are routinely harvested in vegetative phase, as flowering leads to a reduction in stem sugar content (Moore and Nuss 1987). When allowed to reach developmental maturity, sugarcane flowering is highly erratic and the limited resulting seed is short-lived (Rao 1980). Even when produced, seed has low germination rates and yields weak, slow-growing seedlings with low potential for surviving to reproductive maturity. Despite this, *S. officinarum* has been recorded as a minor weed species in some countries (Randall 2002) because it may pose a risk of quarantine disease transmission.

In striking contrast to *S. officinarum*, *S. spontaneum* (L.) is listed as a serious weed in 33 countries. It is adapted to diverse environments throughout tropical and subtropical regions worldwide and most prevalent in central and southeastern Asia (Holm et al. 1997). *S. spontaneum* is further classified as a noxious weed in 46 states of the USA (USDA 2010). Polyploidy likely contributes to *S. spontaneum*'s high phenotypic plasticity and effective competition in disturbed and variable habitats. *S. spontaneum* tolerates a broad range of soil types and moisture levels, is shade tolerant, and produces many small wind-dispersed seeds (Pursglove 1972). *S. spontaneum*'s competitive advantage is also at least in part due to its diverse reproductive abilities. *S. spontaneum* can reproduce both vegetatively (root pieces, rhizomes, axillary nodal buds) and via seed. It spreads aggressively above and belowground by extending rhizomes, tillers, and tertiary shoots to form thick clumps (Pursglove 1972). The dense root mat of *S. spontaneum* is largely impenetrable by young seedlings. Weed status restrictions have limited its utilization, but *S. spontaneum* has still been a resource for introgressing cold tolerance, ratooning ability, and biotic stress tolerance into commercial sugarcane hybrids. Perhaps as an indication that a species' value as a crop may sometimes outweigh its risk as a weed, *S. spontaneum* is currently being evaluated as a dedicated energy crop in Europe (Scordia et al. 2010).

There are a number of species within both genus Saccharum (*Saccharum arundinaceum* (Retz.), *Saccharum bengalense* (Retz.), *Saccharum floridulum* ([Labill.] Warb. ex K. Schum. and Lauterb.), *Saccharum narenga* ([Nees ex Steud.] Wall. ex Hack.), *Saccharum procerum* (Roxb.), and *Saccharum ravennae* (L.)) and closely related genera (*Narenga*, *Miscanthus*, and *Erianthus*) listed in "A Global Compendium of Weeds" (Randall 2002). Also being in the collective Saccharum Complex and having a degree of sexual compatibility (Daniels and Roach 1987), these taxa add further to the available germplasm resources for sugarcane improvement. Having been artificially crossed with Saccharum hybrids (Gupta et al. 1978; Grassl 1980), sorghum adds an even more exceptional gene pool for traits such as drought tolerance and large seed size. The offspring from such hybridizations have generally been of low vigor and fertility, but backcrossing to both parents have been achieved (Grassl 1980). There is a limit to the benefit of such backcrossing, however, as Grassl (1980) recorded after the fourth to fifth generation of backcrossing to sorghum the sugarcane chromosomes were eliminated from the intergeneric hybrids. *Imperata cylindrica* is yet another exceptional example of a species of weedy repute with capacity to hybridize with Saccharum (Sreenivasan et al. 1987). The resulting triploid progeny resembled sugarcane and could be self-fertilized to produce F_2 progeny (Daniels and Roach 1987; Sreenivasan et al. 1987).

3.3 Miscanthus spp.

The genus Miscanthus contains several perennial grasses of value towards forage (Ogura et al. 1999), renewable bioproducts (Cappelletto et al. 2000), and recently dedicated energy crops in both Europe (Clifton-Brown et al. 2001) and North America

(Heaton et al. 2004). The most noteworthy, *Miscanthus*×*giganteus* (*M*×*g*), is a triploid derived from hybridization between *Miscanthus sinensis* and *Miscanthus sacchariflorus* that is capable of large biomass yields in temperate latitudes. *M*×*g* is seed sterile but rhizomatous and produces axillary buds at the lowest two to three nodes on each tiller. As a result, *M*×*g* is considered to have invasiveness potential (Raghu et al. 2006). All Miscanthus species produce lightweight seed caryopses capable of extensive wind-borne distribution. This is likely the predominant trait responsible for *M. sinensis* (Lazarides et al. 1997; Meyer and Tchida 1999) and *Miscanthus floridulus* (Randall 2002) being reported as invasive. *M. sacchariflorus* also produces rhizomes and has been recently listed as a noxious weed in one U.S. state (USDA 2010). As in the previous taxa, Miscanthus offers a gradient from which beneficial crops may be developed and invasive derivatives should be avoided.

3.4 Microstegium

The genus Microstegium consists of roughly 40 species worldwide, with Nepalese browntop (*Microstegium vimineum* (Trin.) A. Camus) the most noteworthy example of contrasting crop: weed value. Nepalese browntop is native to Asia and utilized as a natural forage, but intentional breeding efforts have not been recorded. Aside from the correction (Mehrhoff 2000) of one aberrant report of perennialism (Ehrenfeld 1999), *M. vimineum* is an annual species. As the single example of an annual weed in the Saccharinae, its success is largely due to prolific seed production, seed dormancy, and adaptation to understory shade (Cheplick 2010). Its range has steadily expanded since introduction into Tennessee in 1919 (Fairbrothers and Gray 1972) and is currently found in many states in the eastern US (Redman 1995).

4 Genomic Resources

Molecular dissections of Saccharinae species are justifiably being advanced via comparative approaches around the small, annotated genome of sorghum. Its relatively small size, low level of gene duplication, and close evolutionary relationship to valuable target taxa make it an attractive model for the Andropogoneae. Sorghum is much more closely related to maize than rice (~12 mya versus ~42 mya) (Swigonova et al. 2004a, b; Paterson et al. 2004), similarly closely related (~5 mya) (Sobral et al. 1994), retaining gene order (Ming et al. 1998), and to a degree sexually compatible (De Wet et al. 1976) with sugarcane. The capacity to resolve single and multiple whole-genome duplication events since sorghum's divergence with maize (Swigonova et al. 2004b) and Saccharum (Ming et al. 1998) further affords opportunity to investigate the consequences of extensive polyploidization that has occurred throughout the Saccharinae.

An intensive review of sorghum genomic resources is available (Paterson 2008) and thus not included herein. Some discussion of genomic highlights regarding

perennialism and weediness in grasses, however, is warranted. The genetic control of rhizomatousness, in particular, has been partially dissected. In sorghum, a cross between annual *S. bicolor* and perennial *S. propinquum* was utilized to map genes involved in rhizome development (Paterson et al. 1995). A total of nine distinct chromosomal regions on seven of sorghum's ten chromosomes were found to affect at least one rhizome measurement, accounting for between 5 and 13% of the total variation individually. A similar investigation between perennial and annual rice species mapped two major QTLs (Rhz2 and Rhz3) linked to rhizomatousness. The additional finding that the QTLs were dominant suggested that perenniality might be ancestral to annuality (Hu et al. 2003). Interestingly, Rhz2 and Rhz3 mapped to the same chromosomal regions as two sorghum rhizomatousness markers. The synteny was complicated, however, as the two rice genes appeared to interact with additional unidentified genes (Hu et al. 2003) and the corresponding regions in sorghum accounted for only 12% of rhizomatousness and less than 4% of rhizome regrowth (Paterson et al. 1995). In total, chromosomal segments affecting rhizome development were found on 8 of rice's 12 chromosome pairs and 9 of sorghum's 10 chromosome pairs (Hu et al. 2003).

One exceptional correspondence can be carried to maize, in which several QTLs for rhizomatousness were clustered in the region of maize chromosome 4 orthologous to both the Rhz3 region of rice chromosome 3 and a similar rhizome QTL cluster on sorghum chromosome 6 (Westerbergh and Doebley 2004). A major gene for maturity (*Ma1*) also lies in this region, providing support for this trait's importance towards perennialism. Equally informative, stay-green in sorghum has also been mapped to the *Ma1* region (Crasta et al. 1999). This colocalization is consistent with the hypothesis that life history is an expression of programmed senescence in concert with apical determinacy, as meristem activity and the timing of floral induction are putative components of maturity. Combined, these lines of evidence point strongly to this genomic region for further dissection of perennialism. Recent expression studies enriched for rhizome QTL-specific regions in sorghum have identified a subset of candidate genes that includes GAs as key regulators of rhizome development (Jang et al. 2006), somewhat complicated by the implication that many of these genes with rhizome-enriched expression appear also to have multiple functions (Jang et al. 2009).

A second noteworthy correspondence between sorghum, rice, and maize involves tillering. Two QTLs for tillering in maize mapped near *tb1* (teosinte branched 1) on chromosome 1 (Westerbergh and Doebley 2004), as well as orthologous regions on sorghum chromosome 1 (Paterson et al. 1995) and rice chromosome 3 (Hu et al. 2003) also containing tillering QTLs. As a putative transcription factor involved with apical dominance and differences in axillary branching between maize and teosinte (Hubbard et al. 2002), *tb1* is an interesting candidate gene as a component of perennialism.

The above findings bode well regarding potential opportunities to further understanding of perennialism and its underlying genetic components. Application of these and future insights towards development of perennial crops in an environmentally sensitive manner will further advance the prospects of sustainable agriculture.

References

Blumenthal DM, Jordan NR, Svenson EL (2003) Weed control as a rationale for restoration: the example of tallgrass prairie. Conserv Ecol 7(1):6, http://www.ecologyandsociety.org/vol7/iss1/art6/

Brown AHD, Marshall DR (1981) Evolutionary changes accompanying colonization in plants. In: Scudder GG, Reveal JL (eds) "Evolution Today" Proceedings of the second international congress of systematic and evolutionary biology. Hunt Institute for Botanical Documentation, Carnegie-Mellon University, Pittsburgh, PA, pp 351–363

Buckley YM, Downey P, Fowler SV, Hill R, Memmot J, Norambuena H, Pitcairn M, Shaw R, Sheppard AW, Winks C, Wittenberg R, Rees M (2003) Are invasives bigger? A global study of seed size variation in two invasive shrubs. Ecology 84:1434–1440

Bugg RL (1992) Using cover crops to manage arthropods on truck farms. HortScience 27:741–745

Cappelletto P, Mongardini F, Barberi B, Sannibale M, Brizzi M, Pignatelli V (2000) Papermaking pulps from the fibrous fraction of Miscanthus × giganteus. Ind Crops Prod 11(2–3):205–210

Casady AJ, Anderson KL (1952) Hybridization, cytological, and inheritance studies of a sorghum cross—autotetraploid sudangrass × (johnsongrass × 4n sudangrass). Agron J 43:189–194

Cheplick GP (2010) Limits to local spatial spread in a highly invasive annual grass (Microstegium vimineum). Biol Invasions 12:1759–1771

Chiras DD, Reganold JP (2004) Natural resource conservation: management for a sustainable future, 9th edn. Prentice Hall, Upper Saddle River (NJ)

Clifton-Brown JC, Lewandowski I, Andersson B, Basch G, Christian DG, Bonderup-Kjeldsen J, Jørgensen U, Mortensen J, Riche AB, Schwarz KU, Tayebi K, Teixeira F (2001) Performance of 15 Miscanthus genotypes at five sites in Europe. Agron J 93:1013–1019

Crasta OR, Xu WW, Rosenow DT, Mullet J, Nguyen HT (1999) Mapping of post-flowering drought resistance traits in grain sorghum: association between QTLs influencing premature senescence and maturity. Mol Gen Genet 262:579–588

Culman SW, DuPont ST, Glover JD, Buckley DH, Fick GW, Ferris H, Crews TE (2010) Long-term impacts of high-input annual cropping and unfertilized perennial grass production on soil properties and belowground food webs in Kansas, USA. Agric Ecosyst Environ 137:13–24

Daehler CC (1998) The taxonomic distribution of invasive angiosperm plants: ecological insights and comparison to agricultural weeds. Biol Conserv 84:167–180

Daniels J, Roach BT (1987) Taxonomy and evolution. In: Heinz DJ (ed) Sugarcane improvement through breading, vol 11. Elsevier, Amsterdam, Netherlands, pp 7–84

Davis JG, Edye LA (1959) Sorghum almum Parodi, a valuable summer growing grass. J Aust Inst Agric Sci 25:117–127

Daws MI, Hall J, Flynn S, Pritchard HW (2007) Do invasive species have bigger seeds? Evidence from intra- and inter-specific comparisons. S Afr J Bot 73:138–143

De Wet JMJ, Gupta SC, Harlan JR, Grassl CO (1976) Cytogenetics of introgression from Saccharum into Sorghum. Crop Sci 16:568–572

Dinnes DL, Karlen DL, Jaynes DB, Kaspar TC, Hatfield JL, Colvin TS, Cambardella CA (2002) Nitrogen management strategies to reduce nitrate leaching in tile-drained Midwestern soils. Agron J 94:153–171

Ehrenfeld JG (1999) A rhizomatous, perennial form of Microstegium vimineum (Trin.) A. Camus in New Jersey. J Torrey Bot Club 126:352–358

Ellstrand NC, Schierenbeck KA (2000) Hybridization as a stimulus for the evolution of invasiveness in plants? Proc Natl Acad Sci U S A 97:7043–7050

Ellstrand NC, Whitkus R, Rieseberg LH (1996) Distribution of spontaneous plant hybrids. Proc Natl Acad Sci U S A 93:5090–5093

Fairbrothers DE, Gray JR (1972) Microstegium vimineum (Trin.) A. Camus (Gramineae) in the United States. Bull Torrey Bot Club 99:97–100

Gantzer CJ, Anderson SH, Thompson AL, Brown JR (1990) Estimating soil erosion after 100 years of cropping on Sanborn Field. J Soil Water Conserv 45:641–644

Glover J (2005) The necessity and possibility of perennial grain crops. Renew Agr Food Syst 20:1–4
Glover JD, Culman SW, DuPont ST, Broussard W, Young L, Mangan ME, Mail JG, Crews TE, DeHaan LR, Buckley DH, Ferris H, Turner RE, Reynolds HL, Wyse DL (2010) Harvested perennial grasslands provide ecological benchmarks for agricultural sustainability. Agric Ecosyst Environ 137:3–12
Grassl CO (1980) Breeding Andropogoneae at the generic level for biomass. Sugarcane Breeders' Newsletter 43:41–57
Gupta SC, DeWet JMJ, Harlan JR (1978) Morphology of Saccharum-Sorghum hybrid derivatives. Am J Bot 65:936–942
Hadley HH (1953) Cytological relationships between Sorghum vulgare and Sorghum halapense. Agron J 45:139–143
Hadley HH (1958) Chromosome numbers, fertility, and rhizome expression of hybrids between grain sorghum and Johnson grass. Agron J 50:278–282
Hadley HH, Mahan JL (1956) The cytogenetic behavior of the progeny from a backcross (Sorghum vulgare × S. halapense × S. vulgare). Agron J 48:102–106
Heaton E, Voigt T, Long SP (2004) A quantitative review comparing the yields of two candidate C4 perennial biomass crops in relation to nitrogen, temperature, and water. Biomass Bioenergy 27:21–30
Holm LG, Plucknett DL, Pancho JV, Herberger JP (1977) The world's worst weeds: distribution and biology. University Press of Hawaii, Honolulu, p 609
Holm L, Doll J, Holm E, Pancho J, Herberger J (1997) World weeds. Natural histories and distribution. Wiley, USA
Hu FY, Tao DY, Sacks E, Fu BY, Xu P, Li J, Yang Y, McNally K, Khush GS, Paterson AH, Li ZK (2003) Convergent evolution of perenniality in rice and sorghum. Proc Natl Acad Sci U S A 100:4050–4054
Hubbard L, McSteen P, Doebley J, Hake S (2002) Expression patterns and mutant phenotype of teosinte branched1 correlated with growth suppression in maize and teosinte. Genetics 162: 1927–1935
Jang CS, Kamps TL, Skinner DN, Schulze SR, Vencill WK, Paterson AH (2006) Functional classification, genomic organization putatively cis-acting regulatory elements, and relationship to quantitative trait loci, of Sorghum genes with rhizome-enriched expression. Plant Physiol 142:1148–1159
Jang CS, Kamps TL, Tang H, Bowers JE, Lemke C, Paterson AH (2009) Evolutionary fate of rhizome-specific genes in a non-rhizomatous Sorghum genotype. Heredity 102:266–273
Jordan NR, Vatovec CM (2003) Agroecological benefits from weeds. In: Inderjit (ed) Weed ecology and management. Kluwer, Dordrecht, NL, pp 137–158
Lazarides M, Cowley K, Hohnen P (1997) CSIRO handbook of Australian weeds. CSIRO, Canberra, ACT
Mehrhoff LJ (2000) Perennial Microstegium vimineum (Poaceae): an apparent misidentification? J Torrey Bot Soc 127(3):251–254
Meyer MH, Tchida C (1999) Miscanthus Anderss. produces viable seed in four USDA hardiness zones. J Environ Hort 17(3):137–140
Ming R, Liu SC, Lin YR, da Silva J, Wilson W, Braga D, van Deynze A, Wenslaff TF, Wu KK, Moore PH, Burnquist W, Sorrells ME, Irvine JE, Paterson AH (1998) Detailed alignment of Saccharum and Sorghum chromosomes: comparative organization of closely related diploid and polyploid genomes. Genetics 150:1663–1682
Monoghan N (1979) The biology of Johnson grass (Sorghum halapense). Weed Res 19:261–267
Moore PH, Nuss KJ (1987) Flowering and flower synchronization. In: Heinz DJ (ed) Sugarcane improvement through breeding. Elsevier, Amsterdam, pp 273–311
Morin N (1995) Vascular plants of the United States. In: LaRoe ET, Farris GS, Puckett CE, Doran PD, Mac MJ (eds) Our living resources: a report to the nation on the distribution, abundance, and health of US plants, animals, and ecosystems. US Department of the Interior, National Biological Service, Washington, DC, USA, pp 200–205

Morse LE, Kartesz JT, Kutner LS (1995) Native vascular plants. In: LaRoe ET, Farris GS, Puckett CE, Doran PD, Mac MJ (eds) Our living resources: a report to the nation on the distribution, abundance, and health of US plants, animals, and ecosystems. US Department of the Interior, National Biological Service, Washington, DC, USA, pp 205–209

Ogura S, Kosako T, Hayashi Y, Dohi H (1999) Effect of eating mastication on in-vitro ruminal degradability of Zoysia japonica, Miscanthus sinensis and Dactylis glomerata. Grassland Sci 451:92–94

Paterson AH (2008) Genomics of sorghum. Int J Plant Genomics. doi:10.1155/2008/362451

Paterson AH, Schertz KF, Lin YR, Liu SC, Chang YL (1995) The weediness of wild plants: molecular analysis of genes influencing dispersal and persistence of johnsongrass, Sorghum halepense(L.) Pers. Proc Natl Acad Sci U S A 92:6127–6131

Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc Natl Acad Sci U S A 101: 9903–9908

Paustian K, Bergstrom L, Jansson P, Johnson H (1990) Ecosystem dynamics. Ecol Bull 40:153–180

Piper JK, Kulakow PA (1994) Seed yield and biomass allocation in Sorghum bicolor and F1 and backcross generations of S. bicolor × S. halepense hybrids. Can J Bot 72:468–474

Pursglove JW (1972) Tropical crops: monocotyledons. Longman Scientific and Technical, New York

Raghu S, Anderson RC, Daehler CC, Davis AS, Wiedenmann RN, Simberloff D, Mack RN (2006) Adding Biofuels to the invasive species fire? Science Magazine 313:1742

Randall RP (2002) A global compendium of weeds. R.G. & F.J Richardson, Meredith, Victoria

Randall GW, Huggins DR, Russelle MP, Fuchs DJ, Nelson WW, Anderson JL (1997) Nitrate losses through subsurface tile drainage in CRP, alfalfa, and row crop systems. J Environ Qual 26:1240–1247

Rao PS (1980) Fertility, seed storage and seed viability in sugarcane. In: Proceedings of the International Society of Sugar Cane Technologists, pp 1236–1240

Redman DE (1995) Distribution and habitat types for Nepal Microstegium [Microstegium vimineum (Trin.) Camus] in Maryland and the District of Columbia. Castanea 60:270–275

Robertson GP, Paul E, Harwood R (2000) Greenhouse gases in intensiveagriculture: contributions of individual gases to the radiative forcing of the atmosphere. Science 289:1922–1925

Scordia D, Cosentino SL, Jeffries TW (2010) Second generation bioethanol production from Saccharum spontaneum L. ssp. aegyptiacum (Willd.) Hack. Bioresour Technol 101(14):5358–5365

Sobral BWS, Braga DPV, LaHodd ES, Klein P (1994) Phylogenetic analysis of chloroplast restriction enzyme site mutation in the Saccharinae Griseb. subtribe of Andropogoneae Dumort. tribe. Theor Appl Genet 87:843–853

Soltis DE, Soltis PS, Chase MW, Mort ME, Albach DC, Zanis M, Savolainen V, Hahn WH, Hoot SB, Fay MF, Axtell M, Swensen SM, Prince LM, Kress WJ, Nixon KC, Farris JS (2000) Angiosperm phylogeny inferred from a combined data set of 18S rDNA, rbcL and atpB sequences. Bot J Linn Soc 133:381–461

Sreenivasan TV, Ahloowalia BS, Heinz DJ (1987) Cytogenetics. In: Heinz DJ (ed) Sugarcane improvement through breeding. Elsevier, Amsterdam, pp 211–253

Sturz AV, Matheson BG, Arsenault W, Kimpinski J, Christie BR (2001) Weeds as a source of plant growth promoting rhizobacteria in agricultural soils. Can J Microbiol 47:1013–1024

Swift MJ, Anderson JM (1993) Biodiversity and ecosystem function in agricultural systems. In: Schultz ED, Mooney HA (eds) Biodiversity and ecosystem function. Springer, Berlin, Germany

Swigonova Z, Lai J, Ma J, Ramakrishna W, Llaca V, Bennetzen JL, Messing J (2004a) Close split of sorghum and maize genome progenitors. Genome Res 14(10):1916–1923

Swigonova Z, Lai JS, Ma JX, Ramakrishna W, Llaca M, Bennetzen JL, Messing J (2004b) On the tetraploid origin of the maize genome. Comp Funct Genomics 5:281–284

Thomas H, Thomas HM, Ougham H (2000) Annuality, perenniality and cell death. J Exp Bot 51: 1781–1788

Tilman D, Cassman KG, Matson PA, Naylor R, Polasky S (2002) Agricultural sustainability and intensive production practices. Nature 418:671–677

USDA (2010) Saccharum spontaneum L. Wild Sugarcane. USDA, http://plants.usda.gov/java/profile?symbol=SASP
Warwick SI, Thompson BK, Black LD (1984) Population variation in Sorghum halepense, Johnson grass, at the northern limits of its range. Can J Bot 62:1781–1790
Warwick SI, Phillips D, Andrews C (1986) Rhizome depth: the critical factor in winter survival of Sorghum halepense (L.) Pers. (Johnson grass). Weed Res 26:381–387
Weber E (2003) Invasive plant species of the world—a reference guide to environmental weeds. CABI Publishing, Oxfordshire, UK
Westerbergh A, Doebley J (2004) Quantitative trait loci controlling phenotypes related to the perennial versus annual habit in wild relatives of maize. Theor Appl Genet 109:1544–1553
Williamson M, Fitter A (1996) The varying success of invaders. Ecology 77:1661–1666
Yudelman M, Ratta A, Nygaard D (1998) Pest management and food production looking for the future. Food, Agriculture, and the Environment Discussion Paper 25. International Food Policy Research Institute (IFPRI). Washington, DC

Part V
Synthetic and Futuristic Perspectives

Chapter 22
Bringing the Benefits of Sorghum Genomics to Africa

Segenet Kelemu, Brhane Gebrekidan, and Jagger Harvey

Abstract Sorghum was the first indigenous African crop to have a completed genome sequence. This provides an invaluable tool for understanding sorghum traits genetically and identifying useful molecular markers. The challenge is to extract relevant information to improve the diverse complex traits of sorghum including drought tolerance, disease and pest resistance, and overall yield. The promise of sorghum to improve food security and help lift millions of Africans out of poverty provides a moral imperative for investing in its improvement. Therefore, we must critically and objectively assess all available resources and coordinate our efforts for the benefit of Africans.

Keywords Sorghum • Genomics • Africa • National Agricultural Research Systems (NARS) • Crop improvement • Biotic stress • Abiotic stress • Molecular breeding • Capacity building

1 Background

Sorghum in Africa generally thrives well under a wide range of growing conditions, being dominant in the semiarid tropical belt of the continent where limited rainfall and high temperatures prevail. The altitude range for the crop stretches from near sea level to about 2,500 m, which means it thrives well under hot and dry conditions as well as in very high and cold highlands, such as in Ethiopia, Rwanda, and Uganda.

S. Kelemu (✉) • J. Harvey
Biosciences eastern and central Africa-International Livestock Research Institute Hub (BecA-ILRI Hub), Kabete Campus, Old Naivasha Road, Nairobi 00100, Kenya
e-mail: s.kelemu@cgiar.org

B. Gebrekidan
Agriculture Working Group, Ethiopian Academy of Sciences, Addis Ababa, Ethiopia

A.H. Paterson (ed.), *Genomics of the Saccharinae*, Plant Genetics and Genomics: Crops and Models 11, DOI 10.1007/978-1-4419-5947-8_22,

Historically, the sorghum yield levels in Africa have been very low and have remained stagnant (FAO and ICRISAT 1996; http://faostat.fao.org, based on 1961–2008 data for sorghum yield in Africa). However, the genetic yield potential of the crop could be as high as wheat and maize if the necessary inputs of improved seed, fertilizer, water, crop management, and protection are provided.

The low yield trends in the continent are a major cause for concern, indicating that significant research efforts in crop improvement and management are needed. The primary reasons for the low yields are low genetic potentials of cultivars, disease and insect attack, devastation by birds, shortage of soil moisture, poor soil fertility, and overall poor crop management.

Throughout Africa, sorghum has multiple uses ranging from staple food to construction material. The main traditional foods made from the sorghum grain are leavened breads (such as *injera* in Ethiopia and Somalia and *kisra* in the Sudan), pop-sorghum (*fendisha*), thin and stiff porridges (*ugali*, *genfo*, *tuwo*), fermented and unfermented traditional drinks, as well as many other dishes. The grain, green stalks, and leaves are excellent animal feed, while the dried stalks, particularly of the tall varieties, are essential construction and fuel materials. Many homes and compound fences in sorghum-growing parts of Africa rely heavily on the crop to meet their basic construction material needs. The multitude of uses and broad adaptation of sorghum make it a desired crop on which many African farmers rely for their livelihoods (Gebrekidan and Gebre-Hiwot 1981).

2 African Sorghum Genetic Resources: The Promise of Genomics

Sorghum is indigenous to Africa, with the Northeastern quadrant of the continent recognized as its primary origin (Doggett 1988). The genetic diversity of the crop across the continent is vast; almost any desired trait can be identified from the rich germplasm collection conserved in the custody of the International Crops Research Institute for Semi-Arid Tropics (ICRISAT). The Ethiopian Institute of Biodiversity Conservation (EIBC) also maintains a rich collection of indigenous Ethiopian sorghums.

The five primary races of the crop—Bicolor, Guinea, Kafir, Caudatum, and Durra, along with the ten possible hybrid combinations of these races, forming intermediate races—are all found in different parts of the continent. The distribution of each sorghum race is restricted to specific parts of the continent: Bicolor dominates west and southwest Ethiopia, Tanzania, and Mozambique; Guinea covers southwest Ethiopia, Tanzania, Mozambique, and West Africa; Kafir mostly occurs in Southern Africa; Caudatum is found in Eastern and Central Africa; and Durra is restricted to Ethiopia and Sudan. The specific intermediate races are found in the general geographic areas where their parental primary races overlap (Gebrekidan 1979 and 2009).

The wide genetic variability of sorghum in Africa is still an underexploited resource. Sorghum improvement to date has relied mostly on conventional methods of crop improvement and selection among landraces. There are still unrealized opportunities to utilize African sorghum genetic resources and conventional breeding strategies to improve the yield and stability levels of sorghum. The contributions of genomics research to sorghum improvement in Africa so far have been minimal, excepting the significant progress on *Striga* resistance breeding (Ejeta and Gressel 2007). The potential for genomics in sorghum improvement has just begun to be tapped. Increasing availability of biotechnology tools and Africa's growing capacity to utilize them promise to help break the long trend of flat sorghum yields in the coming years.

3 Constraints to Sorghum Production in Africa

A wide spectrum of constraints limits higher sorghum yields in Africa, among which are biotic (diseases, insects, weeds, and birds) and abiotic (low soil fertility, drought, salinity, soil acidity, cold, and heat) stresses. Economically important sorghum diseases predominantly affect the aerial portions of the plant. Some of the most significant on the continent are anthracnose (*Colletotrichum graminicola*), grey leaf spot (*Cercospora sorghi*), rust (*Puccinia purpurea*), bacterial stripe (*Pseudomonas andropogonis*), charcoal rot (*Macrophomina phaseolina*), covered (*Sphacelotheca sorghi)*, long (*Tolyposporium ehrenbergii)*, and loose (*Sphacelotheca cruenta)* smuts, and molds.

The major insects attacking sorghum in Africa are shoot fly (*Atherigona soccata*), midge (*Contarinia sorghicola*), and several stem borers (*Chilo partellus, Busseola fusca, Sesamia cretica*). The most troublesome insects of sorghum in Africa are the stem borers.

One of the economically important weeds attacking sorghum in the continent is the parasitic weed *Striga,* which can be devastating. Two species of *Striga*, *S. hermonthica* and *S. asiatica*, are widespread in Africa and often destroy entire fields in heavily infested areas. These infested fields are often abandoned as not useful for sorghum production due to the decades of longevity of the *Striga* seeds in the sick soils.

Birds, particularly *Quelea quelea,* are a long standing biotic and social constraint for farmers growing good-quality sorghums with low tannin content. In heavily bird-infested zones, farmers are often forced to plant varieties with high tannin content, which are not liked by either birds or people. The high tannin and bitter grains of brown sorghums make low-quality foods and also have poor digestibility. When an African farmer chooses to grow high-quality sorghum, scouting and chasing away birds is perhaps the single most expensive labor cost. Very often, engagement in chasing away birds from sorghum fields accounts for millions of African children not attending schools.

Among the abiotic stresses, drought is by far the single most important limiting factor in realizing higher sorghum productivity. Although sorghum is recognized as a drought resistant crop, moisture stress is often one of the most important yield-limiting constraints to production in Africa. Varieties can be and have been developed to do well under limited moisture conditions. Early maturing varieties selected and developed for specific agroecologies could help address the moisture shortage problem, but often reduce yield potential.

If moisture is not limiting, sorghum can be grown successfully under a wide range of conditions. Where nitrogen is deficient in the soil, rotation and/or intercropping with a leguminous crop, such as cowpea and beans, is important and is practiced by many African farmers. In general, specific genotypes can be selected to suit nutrient levels of the soils in a given area. If the use of fertilizers is economical in a given situation, one must ensure that this is done under adequate moisture. There is potential, as in the case of maize, to develop nitrogen use-efficient sorghums.

In tackling a number of the constraints listed above, selection from indigenous germplasm resources and conventional breeding have achieved varying degrees of success. In these efforts, national, regional, and international sorghum improvement programs have worked collaboratively to develop cultivars suitable for various agroecological conditions of Africa (Gebrekidan 1987). The appropriate deployment of genomics tools will further enhance the efficiency and effectiveness of sorghum improvement.

4 Agronomic Traits for Targeted Improvement

Plant breeders use conventional methods to genetically modify plants by combining genes for resistance to biotic and abiotic stress factors, crop yield, quality, seed characteristics, and many other traits of agronomic importance. Examples of sorghum cultivars released by national programs following conventional breeding are available with each African national program. The accelerated use of genomics tools in sorghum breeding in Africa may enhance the effectiveness of selection for many economically important traits.

The architecture and overall type of the sorghum plant are critical in positively affecting grain yield and other uses of the crop through a given breeding program. Among the traits that are amenable to genetic improvement through breeding are maturity, plant height, photoperiod sensitivity, panicle type, tiller number, and seed characteristics (size, number, weight, hardness, and color). Modifications of plant height and maturity time are among the main traits breeders manipulate to improve level and sustainability of yield. Maturity time depends on the specific agroecology and ranges from 3 to 9 months or more. Plant height ranges from half a meter to over 4 m. Crop uses (such as forage, construction, fuel), sugar content, and grain type dictate the selection of height group targeted in a breeding project. Another key characteristic is that sorghum can be photoperiod sensitive or insensitive. Panicle characteristics also vary widely, from loose and open forms to compact with erect

to drooping characteristics. Features of interest of grains on the panicle include seed size, number, hardness, color, and weight. Tiller numbers can range from single to multiple depending on the genotype and environmental conditions. With proper management, successive harvests from the same planting are possible due to the ratooning nature of the crop. Therefore, the genetic and phenotypic diversities for these characteristics are readily available for more extensive exploitation in improvement efforts.

Well-established and adept African sorghum breeding programs are already in existence, ready to benefit from an expanded array of synergistic partnerships and technologies. Improvement efforts are currently conducted at the national, regional, and international levels. Historically, at the national level, the strongest programs in Africa have been those of Sudan, Ethiopia, Uganda, Kenya, and Tanzania in the east and Nigeria, Senegal, Mali, Burkina Faso, and Niger in the west. In the last 50 years or so, each of these national programs has developed and released cultivars suitable for its own special conditions. There have also been regional collaborative activities in parts of Eastern, Southern, and West Africa. Internationally, the most active programs have been and continue to be implemented by ICRISAT, the Sorghum, Millet and Other Grains Collaborative Research Support Program (INTSORMIL CRSP), and *icipe*-African Insect Science for Food and Health. Each of these programs has contributed significantly to sorghum improvement in the continent.

Global achievements of conventional sorghum breeding are very strong, correlating with historical application of green revolution technology. With the distinct exception of Africa, the most effective improvements have been achieved in the United States, Mexico, Argentina, and India. The most important conventional science-based technology that has driven their high increases in sorghum production has been the deployment of the hybrid sorghum technology and establishment of associated seed industry. African countries currently lag far behind in development and deployment of hybrid sorghum technology. The lack of strong hybrid sorghum breeding strategy in the national programs and the absence of industries for hybrid seed production are among the major bottlenecks limiting the adoption of hybrid sorghum in Africa.

In strategically selected African countries, well-tested and suitable sorghum hybrids can contribute significantly to increasing food production and improving food security. A hybrid seed development program can also serve as a good platform for stacking additional desirable traits on the parental lines of high-yielding and adapted cultivars.

In Eastern Africa, Sudan is the only country to develop and commercialize a hybrid sorghum, Hagin Dura-1, which is still under production. Achieved with the support of ICRISAT, this serves as an excellent example of effective collaboration of a national program and an international center. In West Africa, the Niger national program has also developed and released a sorghum hybrid, but has not been successful in distributing it widely to a significant number of growers. Potentially suitable A, B, R lines, which are conventionally required for hybrid sorghum seed production, are available to those national programs that have interest in developing and using hybrids, and organizations such as ICRISAT and INTSORMIL CRSP can

play an important role in facilitating their development. Biotechnologies such as marker-assisted selection (MAS) can also play an important role in enhancing the genetic worth of A, B, R parental lines of hybrids through gene pyramiding for complex traits such as yield enhancement, resistance to drought, insects, pathogens, and *Striga* and nutritional quality improvement. In such situations, collaborative work of national, regional, and international programs can effectively exploit the potentials of genomics in sorghum breeding in general and hybrid sorghum development in particular.

African national programs can draw from a wealth of biodiversity, potential partners, and available biotechnologies to expand their crop improvement efforts. In view of the broad sorghum genetic diversity available in and for the various sorghum ecological zones of Africa, it seems practical to give high priority to conventional breeding and selection approaches. At the same time, biotechnology and conventional approaches should be strategically used in concert to help improve overall sorghum production in Africa. Despite the high cost and the associated skilled human power requirements, MAS has been successful in maize in developing high-value lines (Ragot and Lee 2007). A similar success story in sorghum cultivar development has been reported in the area of *Striga* resistance (Grenier et al. 2007; Ejeta and Gressel 2007). Such approaches should obviously be supported by improved agronomic practices and inputs, integrated crop and pest management, improved markets, and stronger research and extension institutions to make significant contributions to sorghum production and utilization across the continent.

5 Constraints to Applications of Genomics in the African Context

Modern biotechnologies have been successfully applied in many countries to tackle constraints that seemed intractable more than a decade ago. The application of molecular biology tools has greatly enhanced our ability to diagnose diseases, determine the genetic diversity of pest and pathogen populations, understand host–pathogen interactions, combine genes from a wide array of organisms in transgenic technologies, and develop durable strategies for their management. Opportunities linking new advances in biosciences to initiatives aiming to solve some of Africa's major agricultural problems remain largely untapped. Similar to successes on other continents, biosciences tools provide powerful means for tackling Africa's agricultural challenges such as pests and diseases, drought and other climatic constraints, and low crop productivity.

Because Africa depends heavily on agriculture to support its growing population, it stands to benefit from technologies that can increase crop productivity, enhance nutritional quality, improve soil fertility, and minimize forest destruction (Kelemu et al. 2003). African farmers encounter several significant challenges ranging from lack of access to technology and inputs to market access. A general focus on short-term objectives, inadequate investment in research, fragmented efforts, and

search for quick and simple solutions to complex problems contributes to the lack of progress in agricultural development. Many African countries are still taking this short-term, quick-fix approach to human resource development as well as infrastructure. The low and declining public expenditure on research and development and the poorly developed or nonexistent private sector in many African countries represent some of the major constraints on science and technology and product development.

Major constraints to biotechnology development and applications in Africa can broadly be grouped into two categories: (a) scientific and technical and (b) limited capacity—human, infrastructure, policy, public awareness, and funding (Mugabe and Ambali 2006). Current specific constraints to the optimal use of modern biosciences in Africa are:

- Lack of sufficient, suitably trained scientists, who are able to follow a career path as scientists in Africa
- Limited involvement of women and girls in science and technology at all levels
- Lack of adequate laboratory facilities and equipment and the human and financial resources essential to operate, maintain, and sustain them
- Limited Internet access and use across Africa
- General lack of enabling environment and subsequent lack of retention of trained human capital in Africa
- Lack of clear problem definition, based on dialogue amongst farmers, consumers, researchers, and the public and private sectors
- Inadequate investment in research and the shortage of operating funds for projects
- Lack of well-developed private sector and limited involvement of this sector in science and technology applications in Africa, in order to develop and deliver products to the market
- Lack of networking within the African scientific community and with the international scientific community to access knowledge and state-of-the-art techniques and deliver potentially useful technologies
- Inadequate characterization, evaluation, and conservation of Africa's rich crop, animal, and soil microbial resources
- Inadequate policies and legal frameworks (e.g., biosafety, intellectual property rights)
- Inadequate awareness and official appreciation of the role of biotechnology in research and development

Many of the 47 countries in sub-Saharan Africa are too small to sustain effective agricultural research systems and to have a critical mass of scientists in many areas of biosciences. Because of small country size and perhaps lack of effective and coordinated strategy, agricultural research systems in sub-Saharan Africa are fragmented into nearly 400 distinct research agencies (Pardey et al. 2006). Current levels of spending on agricultural research and development in sub-Saharan Africa are insufficient for sustained growth. Increasing agricultural spending to 10% or more of national budgets brought success in Asia and Latin America. Average rates of

return on investment in agricultural research and extension, for example, have been documented in the range of 35% for sub-Saharan Africa to 50% in Asia in 700 studies (World Bank 2008).

A comprehensive approach that includes capacity building, creation of enabling environments, and sustained investment in research and development is necessary for successfully addressing some of the key constraints.

6 Advances in Sorghum Genomics

In the era of genomics, an array of biotechnology tools are now available that can potentially complement ongoing breeding efforts. The key is to determine which tools afford a comparative advantage for which breeding projects. All molecular applications, including transgenics, use the foundations of conventional breeding at one stage or another.

Plant breeding products, whether produced using conventional or molecular-assisted methods, will only be as good as the phenotyping that was used in their development. Accurate and precise phenotyping is especially important for complex traits such as drought tolerance, wherein fractional contributions of many minor quantitative trait loci (QTL) can be a challenge to identify amidst environment-related variation. To help address these challenges, efforts are underway to improve phenotyping techniques and related technologies, including centralized phenomics facilities for comparison of traits between highly controlled environmental conditions (Gupta et al. 2010; Tester and Langridge 2010). Phenomics is the high-throughput study of phenotypes that lies at the next frontier for phenotyping. Some characteristics can already be easily phenotyped, such as plant height, tillering, maturity, seed size, and color, but others are more complicated, such as yield, nature of reaction to disease and pest attack, and reaction to pathogen strains. An example of a phenotyping advance that could be of use to national program breeders relates to roots. Although roots underlie many agronomically important traits, they can be difficult to phenotype and are often omitted as a measured characteristic. Root phenotyping systems have been developed (e.g., Iyer-Pascuzzi et al. 2010), which rely on plant growth in media or clear pots.

Many breeders in Africa and around the world could benefit by adoption of standardized phenotyping methods and from introduction of newer phenotyping strategies, relevant to traits of interest. Where robust phenotyping is in place, national program breeders are in a suitable position to identify and apply molecular markers that are linked to these traits. Sorghum is the first completed genome for a crop of African origin (Paterson et al. 2009), which should help researchers develop more, better markers and extrapolate biological information from other sequenced monocots and beyond. Molecular markers can provide a useful surrogate for phenotyping (Varshney et al. 2006), with DNA-based markers being the most useful today. Their incorporation into breeding efforts can reduce the population sizes required, increase the number of varieties that can be developed with the same amount of effort, and

even allow improvement of otherwise intractable, genetically complex traits such as drought tolerance. However, the higher cost of QTL mapping is beyond the current means of many national programs without collaboration with institutes that specialize in the application of biotechnology for crop improvement.

Private maize breeding programs now routinely rely on MAS. One of their largest successes has been inbred line conversion for use in hybrid development (Eathington et al. 2007). Such efforts have significantly increased the rate of genetic gain when compared with conventional breeding, warranting the cost of using MAS. Further improvements in MAS are needed to realize the full benefits of this technology and to allow the transfer of proven approaches and protocols to public breeding programs in developing countries (Ragot and Lee 2007). Despite important strides in marker technologies, the use of MAS has stagnated for the improvement of quantitative traits. The use of MAS in plant breeding has continued to increase in the public and private sectors. However, most applications have been restricted to simple monogenic traits. The failure of current MAS to significantly improve polygenic traits has thus far not delivered on the prediction that MAS would reshape plant breeding programs (Heffner et al. 2009).

Genomic selection (GS) uses the complete set of marker data as performance predictors, rather than identifying individual loci as MAS does. GS delivers more accurate predictions with rapid and lower cost gains from breeding (Jannink et al. 2010). The use of doubled haploids (DH) is most useful in GS. DH technology produces inbreds in only two generations, compared to six or more selfing generations conventionally. Effectiveness of selection under DH is enhanced by the higher genetic variance among DH lines (Mayor and Bernardo 2009). GS predicts the breeding values of lines in a population by analyzing their phenotypes and high-density marker scores. GS would dramatically change the role of phenotyping, which would then serve to update prediction models and no longer to select lines (Heffner et al. 2009). Similar use of the sorghum genome would be useful for identification of a single nucleotide polymorphism (SNP) panel for these genome-wide association studies, as in maize (Gore et al. 2009; discussed later).

The completion of the sorghum genome (Paterson et al. 2009) provides an invaluable tool for help in understanding sorghum traits genetically and allows identification of many additional molecular markers. From the reference genome, Paterson et al. (2009) found ~71,000 simple sequence repeats (SSRs) and identified conserved-intron scanning primers for 6,760 genes, serving as informative molecular markers across monocotyledons, especially for orphan cereals. Seven percent of sorghum's genes are sorghum specific. The challenge now is to mine the information out of them to uncover how sorghum is such a drought-tolerant crop which can grow over many agroecological zones and produce good yield over a wide range of conditions.

With the reference genome for sorghum now available, next-generation sequencing can be easily applied for less expensive whole genome re-sequencing (Ossowski et al. 2008) of important sorghum varieties and for conducting gene expression analysis (Weber et al. 2007; Cheung et al. 2006). Sequencing technologies are becoming increasingly affordable, faster, and more accurate. Solexa/Illumina (http://www.illumina.com/), Applied Biosystems SOLiD (http://www.appliedbiosystems.

com/absite/us/en/home/applications-technologies/solid-next-generation-sequencing.html), and Roche/454 (http://www.genome-sequencing.com/) sequencing are already commonly in use. Smaller scale machines and newer platforms such as Ion Torrent are further reducing costs, making next generation sequencing increasingly more accessible and affordable. It is predicted that these technologies will eclipse high-throughput SNP genotyping platforms such as Illumina Goldengate/BeadXpress once multiple whole genome re-sequencing becomes less expensive.

The real contribution that these available genome sequences and related sequencing technologies can make to national and regional breeding programs is by providing genomic information that facilitates development of better genetic maps and easy-to-use molecular markers. Genotyping tools such as agarose gel-based SSR and SNP genotyping are more accessible and more affordable than high-tech genotyping tools. African national programs and universities are building capacity to use these types of genotyping tools, with the help of various partners and initiatives that are discussed later. Alternatively, national program breeders can choose to integrate genotyping into their breeding efforts by taking advantage of genotyping service provision programs. Whether they use molecular markers directly or through service provision, these biotechnologies can help African breeders more efficiently develop the wide diversity of improved farmer-preferred sorghum varieties required for the agro-ecologically and socially diverse African continent. However, the main bottlenecks for African sorghum breeding programs in taking full advantage of the emerging genomic technologies are still the limited availability of trained personnel coupled with inadequate infrastructure and funding.

Beyond molecular markers, a number of other biotechnologies are available that can capitalize on genomic information for sorghum improvement. One such technology is genetic transformation. Due in large part to difficulty associated with producing transformed plants in vitro, genetic transformation of sorghum may be more difficult than that of many other plant species. Nonetheless, good transformation and regeneration systems have been developed that allow the use of transformation as a tool in sorghum improvement. Production of transgenic plants has been achieved with biolistic transformation of various tissue types, including immature embryos (Casas et al. 1993), immature inflorescences (Casas et al. 1997), and shoot apices. The latter was used to produce transgenic sorghum expressing cry1Ac at low levels to confer partial tolerance to the spotted stem borer (*Chilo partellus* Swinhoe) first instar larvae (Girijashankar et al. 2005), demonstrating that focused efforts should be able to achieve *Bacillus thuringiensis* (Bt) gene-mediated resistance to stem borers in transgenic sorghum. *Agrobacterium*-mediated transformation protocols have also been developed that produce whole transgenic plants (Howe et al. 2006). The African Biofortified Sorghum (ABS) project used *Agrobacterium*-mediated transformation to produce nutritionally enhanced transgenic lines that have undergone field trials in Puerto Rico (Africa Harvest Biotech Foundation International 2007; http://www.supersorghum.org).

Knockdowns of gene expression are a key tool for gene discovery and validation of gene function. As expected from the widely conserved nature of the RNA silencing pathways across plant species, RNA interference (RNAi) is an effective means of downregulating target genes in transgenic sorghum (Cook et al. 2010). Virus-induced gene silencing (VIGS) would also be useful. For example, barley stripe mosaic virus has been used as a VIGS tool in other monocots such as barley. A similar tool would be very useful for high-throughput reverse genetic screens in sorghum. RNAi can not only be used for gene discovery/validation, but also to engineer resistance to viruses.

An important consideration for the application of genetic transformation for sorghum in Africa is introduction into its center of diversity. Studies like the ABS project are currently aiming to assess the potential impact of various genes in wild sorghum (Africa Harvest Biotech Foundation International 2007; http://www.supersorghum.org). Policy development, public awareness, and acceptance of transgenic crops remain the key to deployment of transgenic crop varieties in many African countries.

As discussed earlier, the wide range of sorghum genetic diversity is an invaluable resource for breeding efforts. When the gene(s) underpinning a given trait is known, Targeting Induced Local Lesions In Genomes (TILLING) technology (Till et al. 2007) is a tool that can be used to identify useful genetic variation, either naturally available in germplasm (ecoTILLING) or new variation present in mutagenized populations. Applications of TILLING extend to any trait for which the genes which control it are known. EcoTILLING could help identify useful genetic variation in the vast sorghum collections, making these resources more accessible to breeders.

As with any technology, the comparative advantages of these specific technologies and their applicability to sorghum improvement should be considered on a per breeding project basis. As various laboratories around the world continue to apply these technologies to sorghum, a watchful eye should be cast for any findings relevant to African sorghum improvement.

7 Genomic Lessons for Sorghum from Other Cereals

For crops overall, maize research has played a key role in helping to drive the development and advancement of quantitative genetics since the study of hybrids in the 1950s (Hallauer 2007) and continues to help drive it in the era of molecular breeding. In a recent example, the Buckler laboratory used next-generation re-sequencing of 27 diverse maize inbred lines to identify 2.91 million SNPs for genomewide association studies (Gore et al. 2009). Many lessons from and tools developed for these maize genomewide selection studies (e.g., data analysis methodology and software) can be easily translated for work in sorghum, given that its reference genome is also available.

The rice genome has been available since 2002 (Yu et al. 2002; Goff et al. 2002) and MAS has been applied for the development of improved rice varieties.

For example, the *Sub1* rice QTL contains the *Sub1A* gene, an ethylene response factor-like gene which confers tolerance to submergence during flooding (Xu et al. 2006). From this genomics-based advance, MAS has been used to introgress *Sub1* into mega varieties, producing dramatically submergence-tolerant varieties (Septiningsih et al. 2009). This is an important advance for the billions of people in Asia whose rice staple is currently under threat by potentially wetter conditions associated with climate change. Similar breakthroughs in our understanding of agronomically important cereal genetic pathways could potentially hold the key to radical advances in sorghum yields, especially in the face of climate change.

Comparative genomics, or comparisons between genome sequences of related species, can reveal a great deal about the genetics between shared and contrasting traits (Gale and Devos 1998; Bennetzen 2002; Liang et al. 2008). Comparisons between the sorghum (Paterson et al. 2009), maize (Schnable et al. 2009), rice (Yu et al. 2002; Goff et al. 2002), and Brachypodium (The International Brachypodium Initiative 2010) genomes provide a valuable resource for comparative genomics. Comparing the genomes of these four grasses can provide valuable information about the genetic basis of similarities and differences in agronomic traits between these and related species. For example, the genetic underpinning of sorghum's drought tolerance likely resides in the 7% of its genes that are unique when compared to *Arabidopsis*, rice, and poplar (Paterson et al. 2009). When combined with experimental results, these 7% of genes can quickly yield a small number of candidates that may be involved in a phenotype such as drought tolerance: sorghum has five homologues of rice miRNA 169 g, which is predicted to target a gene linked to drought resistance in maize and *Arabidopsis*: nuclear factor Y (NF-Y) B transcription factor (Paterson et al. 2009; Nelson et al. 2007). Identification of a gene underlying a QTL in one sequenced grass genome may help in immediate identification of a candidate for the same trait in another through genome comparison.

Genetic transformation has also played a role in cereal improvement mainly in maize and rice. Both insect- and herbicide-resistant biotech maize are well established internationally (James 2010). SmartStax, a novel biotech maize containing eight different herbicide tolerance and insect-protection genes in top-performing hybrids, was approved in 2009 for production in North America (http://www.dowagro.com/science/product_updates/smartstax.htm). According to the International Service for the Acquisition of Agri-Biotech Applications (ISAAA), drought-tolerant maize is expected to be deployed in the United States in 2012 and sub-Saharan Africa in 2017. China has developed and approved Bt rice and phytase maize in November 2009 (http://www.isaaa.org/resources/publications/briefs/41/executivesummary/default.asp). With an available reference genome and other biotechnology tools, sorghum is poised to benefit from similar improvement efforts that have already borne fruit and provided relevant lessons in related cereal crops.

There are also lessons from sorghum that will no doubt be invaluable to breeders of maize, rice, and other cereals. Some of sorghum's characteristics that might be transferable to other cereals include its drought tolerance, more efficient C4 photosynthetic system, and perenniality (Paterson et al. 2009). The C4 rice project (http://beta.irri.org/projects15/c4rice) is a consortium led by the International Rice Research Institute (IRRI) and funded by the Bill & Melinda Gates Foundation (BMGF).

Their research aims to transform rice from a C3 to a C4 plant by investigating the basis of C4 photosynthesis and essentially transfer it from maize (C4) into rice. As a second sequenced C4 grass, the sorghum genome could provide valuable insights to their efforts. In other efforts, scientists around the world are working on various strategies to develop drought-tolerant and other climate-resilient crops. Although challenging, drought resistance can be improved through conventional breeding, using existing genetic diversity. Modern tools, involving molecular markers, genetic engineering, and comprehensive gene expression profiling, provide opportunities for directing the continued breeding of genotypes that provide stable grain yield under widely varied environmental conditions (Bruce et al. 2002).

8 Marker-Assisted Selection in African Sorghum Breeding

Striga resistance breeding has progressed further in sorghum than any other crop. The importance of molecular markers for enhancing efficiency of breeding for *Striga* resistance in sorghum via gene mapping and MAS has been effectively demonstrated (Ejeta and Gressel 2007; Grenier et al. 2007). This is largely because screening for complex resistance under field conditions is difficult and often unreliable. Since *Striga* resistance involves a complex mechanism, the use of bioassays to dissect it into simply inherited components (e.g., germination, haustorial initiation, hypersensitive response) along with genetic analysis and molecular mapping of these simple traits was essential. Marker-mediated gene pyramiding, the stacking of several genes controlling each component of resistance into a single genotype, was successfully achieved in the variety Brhan developed from the cross between the *Striga*-resistant varieties SRN39 and Framida. Brhan combines low germination stimulant, hypersensitive response, and incompatible reaction with high yield in the Wollo region of Ethiopia where it was released and is being disseminated (Grenier et al. 2007; Ejeta and Gressel 2007).

Another sorghum trait that researchers are attempting to improve using MAS is drought tolerance, or the "stay-green" trait. Stay-green drought tolerance includes improved grain filling and yield during stress (Rosenow and Clark 1981), increased stem carbohydrates (McBee 1984), reduced lodging, and reduced charcoal rot susceptibility (Mughogho and Pande 1984). ICRISAT has been working to map and introgress stay-green QTLs (Haussmann et al. 2002; Mahalakshmi and Bidinger 2002), which are being evaluated in field trials (Santie deVilliers, personal communication).

9 Current Ability of NARS to Benefit from Genomics Research

African National Agricultural Research Systems (NARS) have used molecular tools to complement their breeding efforts sparsely, but to varying degrees. A number of factors have contributed to the limited use of these technologies by NARS,

including cost, expertise, and infrastructure as described earlier. Several initiatives have been actively working to bolster the capacity of NARS breeders and the researchers who aim to support them, as well as provide facilities and expertise open for use in collaborative efforts to improve African crops. Incorporation of molecular markers and other biotechnologies into national program crop improvement efforts assumes availability of expertise and adequate funds. Through various partnerships, there are success stories of incorporation of molecular markers in national program crop improvement, such as *Striga* resistance in sorghum, mentioned earlier. Some considerations and an overview of the use of molecular markers in NARS are discussed below.

Comparative advantage should be considered for a given national breeding program and its projects, whether money would be best spent incorporating molecular markers or improving their conventional breeding program. Some studies have compared the use of conventional breeding versus MAS for improvement of specific traits. Introgression of a single, dominant allele from a donor to a recipient line in maize was considered; while it is more costly to use MAS, it is also quicker (Morris et al. 2003). The authors outlined four considerations for determining comparative advantage of MAS versus conventional breeding for national programs: phenotypic versus marker screening costs; time difference between conventional versus marker-based approach; benefits (time and magnitude) derived from marker strategy-based earlier release; and availability of breeding program funds. Another consideration for propriety of MAS in a breeding project is how readily the characteristic in question can be measured. Of course, these considerations are a moot point if the capacity to apply molecular markers in breeding programs is absent. The application of molecular markers in NARS is becoming increasingly common and possible with the growing body of well-trained breeders. Various efforts, discussed below, are now providing an increasingly solid foundation for use of markers in many national breeding programs.

Molecular markers vary in the cost and level of laboratory expertise required for their use. African national programs have made use of a range of the various molecular markers. The authors of Marker-Assisted Selection, Current Status and Future Perspectives in Crops, Livestock, Forestry and Fish (FAO 2007) conducted a survey of the use of biotechnology in African national programs. They found that the most commonly used molecular markers in crops, in order of most to least frequently used, are random amplification of polymorphic DNA (RAPD), SSRs, amplified fragment length polymorphism (AFLP), and restriction fragment length polymorphism (RFLP), with a category of "other or not specified" likely containing others, perhaps including SNPs. With the advent of modern genomics, this profile will change over time if the research community works together to provide more, easy-to-use markers to national programs, either for their use directly or through service provision centers. For example, beta-carotene is being increased in maize by incorporation of inexpensive polymerase chain reaction (PCR)-based MAS in national program efforts (Yan et al. 2010). The capacity of African NARS to incorporate the use of molecular markers, including collaboration-based efforts with the Consultative

Group for International Agricultural Research (CGIAR) centers and other international and non-African organizations, is discussed in detail later in this chapter.

The use of molecular markers in NARS is on the rise. A recent search of the Food and Agriculture Organization (FAO) Biotechnology in Developing Countries (BioDeC) database in May 2010 retrieved 3,104 crop projects, with 46 projects using biotechnology tools to improve sorghum. Twenty of these involve the use of molecular markers in sorghum in Burundi, Ethiopia, Ghana, Kenya, Mali, South Africa, Sudan, Uganda, and Zimbabwe. The FAO-BioDeC database is not complete and does not include those associated with CGIAR center efforts, but provides a glimpse at current use of biotechnology tools nonetheless. Therefore, African NARS are applying information from sorghum genomics for crop improvement.

The Biosciences eastern and central Africa-International Livestock Research Institute (BecA-ILRI) Hub activities are among several major efforts to boost capacity for NARS to incorporate molecular biotechnology into their agricultural improvement efforts. Hub training courses held in 2009 included 250 participants, with topics including molecular marker-assisted breeding, bioinformatics, data analysis, and sequencing technologies, among others. The MAS-related courses are designed to empower NARS researchers to effectively integrate the use of molecular markers to enhance the efficiency of their crop improvement efforts. They include a heavy emphasis on data analysis and sample preparation, enabling participants to use full genotyping services available at the Hub and elsewhere. Aside from course participants, many scientists and graduate students from across Africa have received short-term training or conducted part of their research at the bench at the Hub, similarly building their capacity to integrate new technologies in their agricultural improvement efforts. Demand for DNA sequencing and genotyping services provided by the Hub has consequently increased over 600% in the last 3 years due to use by researchers from across Africa and internationally.

Other applications of molecular markers are critical to helping breeders tap the diversity resident in genebanks. Diversity studies of sorghum genebanks can make their valuable biodiversity more readily available to breeders. For example, a Generation Challenge Program (GCP)-funded ICRISAT project carried out in part at the BecA-ILRI Hub worked with NARS from Kenya, Uganda, Eritrea, Sudan, Tanzania, Rwanda, Burundi, and Ethiopia to use markers to assess diversity in their germplasm. Coupled with downstream phenotyping, which can be performed on a subset of selected accessions after analysis with markers, genebank genetic gems are becoming more accessible to breeders.

Several of the efforts outlined above offer a model for incorporation of genomic information in national program breeding efforts. National program breeders are best and uniquely suited to lead improvement of African crops for the continent's smallholder farmers. They best know the crop, farmer preferences, constraints, and available germplasm. A number of organizations are now in place to help support them in what they do best. Organizations such as the African Agricultural Technology Foundation (AATF) and ISAAA facilitate knowledge transfer so that proprietary biotechnologies can be incorporated into African crop improvement.

10 Towards Creating an Enabling Environment in Africa

Although there is a general lack of appreciation for the power of science and technology in many African countries, their role in Africa's development has received the increasing attention of the international and continental communities in recent years. The 2003 African Ministerial Conference on Science and Technology, organized by the New Partnership for Africa's Development (NEPAD) Secretariat with the support of the United Nations Education, Scientific and Cultural Organization (UNESCO) and others, established the African Ministerial Council on Science and Technology (AMCOST) and its Steering Committee for Science and Technology as the overall governance structure for setting continental priorities and policies relevant to the development and application of science and technology for Africa's development. The new approach consolidates science and technology programs of the African Union (AU) Commission and NEPAD. The main components of the plan are (a) capacity building, (b) knowledge generation, and (c) technological innovation. The Plan of Action puts emphasis on developing an African system of research and technological innovation by establishing networks of centers of excellence dedicated to specific research and development (R&D) and capacity building programs. Such centers would be networked within the continent as well as promoted to establish links to the international arena. Subsequently four regional biosciences centers were created:

- *BecA* (Biosciences eastern and central Africa) for 18 countries in east and central Africa
- *NABNet* (North African Biosciences Network) for six countries in North Africa
- *SANBio* (Southern African Network for Biosciences) for 12 southern African countries
- *WABNet* (West African Biosciences Network) consisting of 15 countries of the Economic Community of West African States (ECOWAS)

As an example of a center that has become well established, BecA aims to employ modern biotechnology to improve agriculture in eastern and central Africa. It also seeks to strengthen the capacity of scientists in eastern and central Africa to conduct bioscience research and to significantly contribute to improved products that can enhance livelihoods of farmers in the region. The Government of Canada, through the Canadian International Development Agency (CIDA) with contributions by ILRI, funded design and implementation phases of the BecA Project from 2004 to 2009. BecA has a Hub (BecA-ILRI Hub; http://hub.africabiosciences.org/) with a state-of-the-art shared biosciences facility located on the Nairobi, Kenya, campus of ILRI, that provides a research platform, research-related services, and capacity building opportunities for the east and central African countries and beyond. The BecA activities focus on addressing key constraints in African agriculture. In view of the dearth of sufficient expertise in science and technology in Africa, the Hub sees implementing a strong program in capacity building and training as central to achieving its mission. It provides opportunities to increase the capacity of institutions

and individuals to conduct biosciences-related research in Africa and to develop and deliver new technologies.

In addition to the above examples of improved infrastructure for research and capacity building activities, research funding opportunities focused on food security are increasing. For example, the Governments of Australia and China are investing in Africa. Foundations such as BMGF and several other donor agencies and countries are focusing their investments on improving the performance of African agriculture. For example, AATF is leading the Water Efficient Maize for Africa (WEMA) project, a public–private partnership project funded by BMGF and the Howard G. Buffett Foundation. The project aims to develop drought-tolerant maize varieties for Africa using conventional breeding, marker-assisted breeding, and transgenic technologies. AATF works with the International Maize and Wheat Improvement Center (CIMMYT), Monsanto, and the NARS of Kenya, Uganda, Tanzania, South Africa, and Mozambique in this effort. These partners will be contributing their technology, time, and expertise to this 5-year project on developing and field testing African crop varieties.

Another BMGF-funded project named "Harnessing Opportunities for Productivity Enhancement (HOPE) of Sorghum and Millets in sub-Saharan Africa and South Asia" is also underway now. The sorghum portion of HOPE will pursue discoveries and development in the areas of plant breeding and associated improved crop management practices for Africa. The sorghum breeding strategy is aimed at discovering and developing valuable new traits (e.g., yield responsiveness to management practices, resistance to diseases and pests, drought resistance, and grain quality). The approach will integrate conventional with molecular tools to enhance efficiency and precision of the breeding process.

A number of other organizations and initiatives are collectively bolstering capacity for use of biotechnology in African crop improvement efforts. These include programs that support graduate and technical training of African students and scientists in the region and abroad, and set up infrastructure for use of modern biosciences on African soil.

The Association for Strengthening Agricultural Research in Eastern and Central Africa (ASARECA) is another organization that has built capacity for biosciences research in Africa. This regional organization involves the National Agricultural Research Institutes (NARIs) of 11 countries: Burundi, Democratic Republic of Congo, Eritrea, Ethiopia, Kenya, Madagascar, Rwanda, South Sudan, Sudan, Tanzania, and Uganda. ASARECA's services to its partners include using biotechnology as a tool to enhance utilization of biodiversity, including sorghum genetic diversity in the region (http://www.asareca.org).

The Eastern Africa Regional Program and Research Network for Biotechnology, Biosafety and Biotechnology Policy Development (BIO-EARN), funded by the Swedish International Development Agency (SIDA), started in 1999. Over the last 10 years, BIO-EARN supported the training of 31 PhD and 60 MSc graduates in advanced molecular techniques and tissue culture for the partner institutions in Eastern Africa. While the first part of fellows' training was conducted in Sweden, the program then supported their research back at their home institutions in Eastern Africa.

BIO-EARN has now evolved into a new program, Bio-resources Innovations Network for Eastern Africa Development (BIO-INNOVATE) (2010–2014), which is building on the successes of the earlier program. BIO-INNOVATE is colocated on the ILRI campus with the BecA-ILRI Hub.

In all of the above examples, effective partnership between African institutions and their partners is essential to build sustainable human capacity and achieve success in food security for Africa. In this regard, African institutions themselves have to play the lead role in forging the desired and sustainable partnership with the regional and global scientific community.

11 Conclusions and the Way Forward

Currently, sorghum yield levels in general in Africa are very low and have stayed unchanged historically. The low yield trends in the continent are a major cause for concern, indicating that significant research efforts in crop improvement and management are needed to reverse this trend. There are rich opportunities in exploiting African sorghum genetic resources and conventional breeding strategies to improve the yield and stability levels of sorghum. Furthermore, the latest biosciences, including genomics, have a critical role to play in enhancing African sorghum production and food security. The contributions of genomic research in sorghum improvement in Africa so far have been minimal, with the exception of the significant progress on *Striga* resistance breeding. For transgenic crop development and deployment, biosafety regulations are in place in several African countries, including Burkina-Faso, Egypt, South Africa, Kenya, and Uganda. Biotechnology and conventional approaches should be ideally used in concert with each other to improve overall sorghum production in Africa.

Sorghum was the first indigenous African crop to have a completed genome sequence. This provides an invaluable tool in understanding sorghum traits genetically and allows identification of many useful molecular markers. The challenge is to extract the relevant information to tackle the diverse complex traits of sorghum, such as drought tolerance, disease and pest resistance, and overall yield.

The importance of capacity building in sustainable agricultural development cannot be overemphasized. It is reflected in the many initiatives that have been put forward at national, regional, and international levels. Human resources are ultimately a key factor behind any development progress. Some major bottlenecks for effective applications of biotechnologies in Africa comprise shortages of trained personnel, lack of research and development infrastructure, lack of appropriate policies, shortage of sustained funding, and ineffectual means for information dissemination.

Links for capacity building with institutions in developed countries should be used to transfer (when feasible) technology that is available so that this can be efficiently used in African research institutions. The links should also be used to build an essential core of human capacity conversant in the area of biotechnology to represent different areas and needs.

The promise of sorghum to improve food security and help lift millions of Africans out of poverty provides a moral imperative for investing in its improvement. Therefore, we must critically and objectively assess all available resources and coordinate our efforts for the benefit of Africans.

References

Africa Harvest Biotech Foundation International (AHBFI) (2007) A global vision with an African focus to fight poor nutrition with nutrient-rich crops. The Africa Biofortified Sorghum Project: Mid-Term Report, December 2007. Nairobi, Kenya; Johannesburg, South Africa; Washington DC, USA, p 40

Bennetzen J (2002) The rice genome. Opening the door to comparative plant biology. Science 296:60–63

Bruce WB, Edmeades GO, Barker TC (2002) Molecular and physiological approaches to maize improvement for drought tolerance. J Exp Bot 53:13–25

Casas AM, Kononowicz AK, Zehr UB, Tomes DT, Axtell JD, Butler LG, Bressan RA, Hasegawa PM (1993) Transgenic sorghum plants via microprojectile bombardment. Proc Natl Acad Sci U S A 90:11212–11216

Casas AM, Kononowicz AK, Haan TG, Zhang L, Tomes DT, Bressan RA, Hasegawa PM (1997) Transgenic sorghum plants obtained after microprojectile bombardment of immature inflorescences. In Vitro Cell Dev Biol Plant 33:92–100

Cheung F, Haas BJ, Goldberg SM, May GD, Xiao Y, Town CD (2006) Sequencing Medicago truncatula expressed sequenced tags using 454 Life Sciences technology. BMC Genomics 7:272

Cook D, Rimando AM, Clemente TE, Schroder J, Dayan FE, Nanayakkara NP, Pan Z, Noonan BP, Fishbein M, Abe I, Duke SO, Baerson SR (2010) Alkylresorcinol synthases expressed in Sorghum bicolor root hairs play an essential role in the biosynthesis of the allelopathic benzoquinone sorgoleone. Plant Cell 22:867–887

Doggett H (1988) Sorghum, 2nd edn. Longman Scientific and Technical, Harlow, UK

Eathington SR, Crosbie TM, Edwards MD, Reiter RS, Bull JK (2007) Molecular markers in a commercial breeding program. Crop Sci 47(S3):S154–S163

Ejeta G, Gressel J (2007) Integrating new technologies for Striga control: towards ending the witch-hunt. World Scientific, London

Food and Agriculture Organization of the United Nations (2007) Marker-assisted selection—current status and future perspectives in crops, livestock, forestry and fish. FAO, Rome

Food and Agriculture Organization of the United Nations/the International Crops Research Institute for the Semi-Arid Tropics (1996) The world sorghum and millet economies: facts, trends and outlook. FAO/ICRISAT, Rome/India

Gale MD, Devos KM (1998) Plant comparative genetics after 10 years. Science 282:656–659

Gebrekidan B (1979) Sorghum genetic resources in Africa. Eth J Agri Sci 1:108–115

Gebrekidan B (1987) Sorghum improvement and production in Eastern Africa. In: Menyonga JM, Bezuneh T, Youdeowei A (eds) Food grain production in semi-arid Africa. OAU/STRC-SAFGRAD, Ouagadougou, Burkina Faso, pp 141–154

Gebrekidan B (2009) Linking science to product development: focusing on sorghum. Paper presented in the BecA-Syngenta Foundation for Sustainable Agriculture Foundation conference held on April 29, 2009 at the BecA-ILRI Hub, Nairobi, Kenya

Gebrekidan B, Gebre-Hiwot B (1981) Sorghum injera preparations and its quality parameters, Proc. of the Intl. Symp. on Sorghum Grain Quality. ICRISAT, Hyderabad, India

Girijashankar V, Sharma HC, Sharma KK, Swathisree V, Prasad LS, Bhat BV, Royer M, Secundo BS, Narasu ML, Altosaar I, Seetharama N (2005) Development of transgenic sorghum for insect resistance against the spotted stem borer (Chilo partellus). Plant Cell Rep 24:513–522

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchison D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Paszkowski U, Zhang S, Colbert M, Sun WL, Chen L, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu Y, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T, Oliphant A, Briggs S (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). Science 296:92–100

Gore MA, Chia JM, Elshire RJ, Sun Q, Ersoz ES, Hurwitz BL, Peiffer JA, McMullen MD, Grills GS, Ross-Ibarra J, Ware DH, Buckler ES (2009) A first-generation haplotype map of maize. Science 326:1115–1117

Grenier C, Ibrahim Y, Haussmann B, Kiambi D, Ejeta G (2007) Marker-assisted selection for Striga resistance in sorghum. In: Ejeta G, Gressel J (eds) Integrating new technologies for Striga control: towards ending the witch-hunt. World Scientific, London

Gupta PK, Kumar J, Mir RR, Kumar A (2010) Marker-assisted selection as a component of plant breeding. In: Janick J (ed) Plant breeding reviews, vol 33. Wiley, New Jersey, pp 145–205

Hallauer AR (2007) History, contribution, and future of quantitative genetics in plant breeding: lessons from maize. Crop Sci 47(S3):S4–S19

Haussmann BI, Mahalakshmi V, Reddy BV, Seetharama N, Hash CT, Geiger HH (2002) QTL mapping of stay-green in two sorghum recombinant inbred populations. Theor Appl Genet 106:133–142

Heffner EL, Sorrels ME, Jannink J-L (2009) Genomic selection for crop improvement. Crop Sci 49:1–12

Howe A, Sato S, Dweikat I, Fromm M, Clemente T (2006) Rapid and reproducible Agrobacterium-mediated transformation of sorghum. Plant Cell Rep 25:784–791

Iyer-Pascuzzi AS, Symonova O, Mileyko Y, Hao Y, Belcher H, Harer J, Weitz JS, Benfey PN (2010) Imaging and analysis platform for automatic phenotyping and trait ranking of plant root systems. Plant Physiol 152:1148–1157

James C (2010) Global status of commercialized Biotech/GM Crops: 2009, ISAAA Brief 41. ISAAA, Ithaca, New York

Jannink JL, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. Brief Funct Genomic Proteomic 9:166–177

Kelemu S, Mahuku G, Fregene M, Pachico P, Johnson N, Calvert L, Rao I, Buruchara R, Amede T, Kimani P, Kirkby P, Kaaria S, Ampofo K (2003) Harmonizing the agricultural biotechnology debate for the benefit of African farmers. Afr J Biotechnol 2(11):394–416

Liang C, Jaiswal P, Hebbard C, Avraham S, Buckler ES, Casstevens T, Hurwitz B, McCouch S, Ni J, Pujar A, Ravenscroft D, Ren L, Spooner W, Tecle I, Thomason J, Tung CW, Wei X, Yap I, Youens-Clark K, Ware D, Stein L (2008) Gramene: a growing plant comparative genomics resource. Nucleic Acids Res 36:D947–D953

Mahalakshmi V, Bidinger FR (2002) Evaluation of stay-green sorghum germplasm lines at ICRISAT. Crop Sci 42:965–974

Mayor PJ, Bernardo R (2009) Genomewide selection and marker-assisted recurrent selection in double haploid versus F_2 populations. Crop Sci 49:1719–1725

McBee GG (1984) Relation of senescence, nonsenescence, and kernel maturity to carbohydrate metabolism in sorghum. In: Mughogho LK (ed.) Sorghum root and stalk rots: a critical Review. Proc. Consult. Group Discussion on Research Needs and Strategies for Control of Sorghum Root and Stalk Rot Diseases, Bellagio, Italy. 27 Nov–2 Dec 1983. ICRISAT, Patancheru, India, pp 119–129

Morris M, Dreher K, Ribaut J-M, Khairallah M (2003) Money matters (II): costs of maize inbred line conversion schemes at CIMMYT using conventional and marker-assisted selection. Mol Breed 11:235–247

Mugabe J, Ambali A (2006) Africa's science and technology consolidated plan of action. The NEPAD Office of Science and Technology, Pretoria, South Africa

Mughogho LK, Pande S (1984) Charcoal rot of sorghum. In: Mughogho LK (ed) Sorghum root and stalk rots: a critical Review. Proc. Consult. Group discussion on research needs and strategies for control of sorghum root and stalk rot diseases, Bellagio, Italy. 27 Nov–2 Dec 1983. ICRISAT, Patancheru, India, pp 11–24

Nelson DE, Repetti PP, Adams TR, Creelman RA, Wu J, Warner DC, Anstrom DC, Bensen RJ, Castiglioni PP, Donnarummo MG, Hinchey BS, Kumimoto RW, Maszle DR, Canales RD, Krolikowski KA, Dotson SB, Gutterson N, Ratcliffe OJ, Heard JE (2007) Plant nuclear factor Y (NF-Y) B subunits confer drought tolerance and lead to improved corn yields on water-limited acres. Proc Natl Acad Sci U S A 104:16450–16455

Ossowski S, Schneeberger K, Clark RM, Lanz C, Warthmann N, Weigel D (2008) Sequencing of natural strains of Arabidopsis thaliana with short reads. Genome Res 18:2024–2033

Pardey PG, Alston JM, Piggott RR (2006) Agricultural R&D in the developing world: too little, too late? International Food Policy Research Institute, Washington, DC, USA

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboobur R, Ware D, Westhoff P, Mayer KF, Messing J, Rokhsar DS (2009) The Sorghum bicolor genome and the diversification of grasses. Nature 457:551–556

Ragot M, Lee M (2007) Marker-assisted selection in maize: current status, potential, limitations and perspectives from the private and public sectors. In: Guimarães E, Ruane J, Scherf B, Sonnino A, Dargie JD (eds) Marker-assisted selection: current status and future perspectives in crops, livestock, forestry and fish. Food and Agriculture Organization of the United Nations, Rome

Rosenow DT, Clark LE (1981) Drought tolerance in sorghum. In: Loden HD, Wilkinson D (eds) Proc 36th Annu. Corn and Sorghum Industry Res Conf Chicago, IL 9–11 Dec. 1981. Am. Seed Trade Assoc, Washington, DC, pp 18–31

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen W, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He R, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin J, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren L, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Van Buren P, Vaughn MW, Ying K, Yeh CT, Emrich SJ, Jia Y, Kalyanaraman A, Hsia AP, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia JM, Deragon JM, Estill JC, Fu Y, Jeddeloh JA, Han Y, Lee H, Li P, Lisch DR, Liu S, Liu Z, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang L, Yu Y, Zhang L, Zhou S, Zhu Q, Bennetzen JL, Dawe RK, Jiang J, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA, Wilson RK (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115

Septiningsih EM, Pamplona AM, Sanchez DL, Neeraja CN, Vergara GV, Heuer S, Ismail AM, Mackill DJ (2009) Development of submergence-tolerant rice cultivars: the Sub1 locus and beyond. Ann Bot 103:151–160

Tester M, Langridge P (2010) Breeding technologies to increase crop production in a changing world. Science 327:818–822

The International Brachypodium Initiative (2010) Genome sequencing and analysis of the model grass Brachypodium distachyon. Nature 463:763–768

Till BJ, Comai L, Henikoff S (2007) TILLING and EcoTILLING for crop improvement. In: Varshney RK, Tuberosa R (eds) Genomic assisted crop improvement: genomics approaches and platforms. Springer, Dordrecht, The Netherlands, pp 333–349

Varshney RK, Hoisington DA, Tyagi AK (2006) Advances in cereal genomics and applications in crop breeding. Trends Biotechnol 24:490–499

Weber AP, Weber KL, Carr K, Wilkerson C, Ohlrogge JB (2007) Sampling the Arabidopsis transcriptome with massively parallel pyrosequencing. Plant Physiol 144:32–42

World Bank (2008) World development report 2008: agriculture for development. The World Bank, Washington, DC

Xu K, Xu X, Fukao T, Canlas P, Maghirang-Rodriguez R, Heuer S, Ismail AM, Bailey-Serres J, Ronald PC, Mackill DJ (2006) Sub1A is an ethylene-response-factor-like gene that confers submergence tolerance to rice. Nature 442:705–708

Yan J, Kandianis CB, Harjes CE, Bai L, Kim EH, Yang X, Skinner DJ, Fu Z, Mitchell S, Li Q, Fernandez MG, Zaharieva M, Babu R, Fu Y, Palacios N, Li J, Dellapenna D, Brutnell T, Buckler ES, Warburton ML, Rocheford T (2010) Rare genetic variation at Zea mays crtRB1 increases beta-carotene in maize grain. Nat Genet 42:322–327

Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Li J, Liu Z, Qi Q, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Zhao W, Li P, Chen W, Zhang Y, Hu J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Tao M, Zhu L, Yuan L, Yang H (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. indica). Science 296:79–92

Chapter 23
Synthesis: Fundamental Insights and Practical Applications from the Saccharinae Clade

Andrew H. Paterson

Abstract The Saccharinae clade offers both opportunities to improve the efficiency and sustainability at which we convert solar energy and other resources into food, feed, fiber, and fuel and to gain new insights into the ecology, evolution, and function of plant species, their genomes, and their constituent genes. Singular features of biogeography, productivity, and stress tolerance of key Saccharinae taxa fit particularly well with existing or anticipated needs of agriculture. Sorghum holds particular promise as a botanical model for the clade, albeit with more complex genomes in the clade also offering intriguing opportunities to clarify roles of polyploidy in agricultural productivity and post-polyploidy evolution.

Keywords Sorghum • Saccharum • Miscanthus • Karyotype evolution • Carbon assimilation • Drought • Perenniality • Weediness • Invasiveness

1 Background

The Saccharinae, as circumscribed by Kellogg (Chap. 1, this volume), has a rich history of contributions to humanity with the promise of still-greater contributions as a result of recent invigorated interest and research activity in several members of this group. Both sorghum (Kimber et al., Chap. 2) and Saccharum (Paterson et al., Chap. 3) appear to have been of importance even to prehistoric peoples. Scientific improvement through prudent engagement of conventional approaches and genetic engineering (Tejinder et al., Chap. 10; Beyene et al., Chap. 11) now offer the potential to transcend the limits of natural variation in enhancing the productivity, quality,

A.H. Paterson (✉)
Plant Genome Mapping Laboratory, Departments of Crop and Soil Science, Plant Biology, and Genetics, University of Georgia, Athens, GA, USA
e-mail: paterson@plantbio.uga.edu; paterson@uga.edu

A.H. Paterson (ed.), *Genomics of the Saccharinae*, Plant Genetics and Genomics: Crops and Models 11, DOI 10.1007/978-1-4419-5947-8_23,

and sustainability of conventional economic products from these plants, and further to harness these plants as cost-effective biofactories for production of novel high-value products. Miscanthus, although less familiar as a crop and indeed arguably not yet domesticated, also has a long history of human utilization (Sacks et al., Chap. 4), with recent interest in its tremendous biomass productivity motivating the development of tools and technologies to expedite its improvement (e.g., Engler et al., Chap. 12). Parallels in gene repertoire, organization, and function among these taxa (Paterson et al., Chap. 18) suggest that knowledge of the composition (Vermerris, Chap. 17; Murray, Chap. 20), agronomy (Jessup, Chap. 21), pathology (Magill, Chap. 15), and entomology (Huang et al, Chap. 16) of one or more of the Saccharinae crops may "translate" well to other members of the group, conferring synergies in the parallel improvement of these taxa.

By way of summary and synthesis, here I suggest a few areas in which the Saccharinae promise to make singularly large contributions to plant biology generally, and in particular to practical application of plant biology in the production of food, feed, forage, and fiber.

2 Functional and Evolutionary Genomics

About 20 million years ago (mya), a population of $2n = 2x = 20$ grassy plants had a singularly bright future. An early branch diverged from this population ~10–15 mya, experiencing genome-wide duplication and genome size expansion accompanied by chromosome number reduction leading to the modern maize genome (Gaut et al. 1997; Swigonova et al. 2004a). About 8–9 mya (Sobral et al. 1994; Jannoo et al. 2007) another branch diverged, also experiencing genome-wide duplication and leading to the modern genomes of Saccharum, perhaps the most important biofuel crop worldwide, and its close relative *Miscanthus* that is among the highest-yielding herbaceous temperate biomass crops known (Heaton et al. 2008). A third branch spawned lineages with a wide range of fates (Spangler et al. 1999), one being domesticated into cultivated sorghum, and another experiencing polyploidy to form one of the world's most aggressive weeds (*Sorghum halepense*).

Among these important grasses, the Andropogoneae, the genome of *Sorghum bicolor* L. Moench. (sorghum), remains the best representative known of the common ancestor of ~20 mya. The sorghum karyotype is suggested by parsimony to have been shared by the Andropogoneae common ancestor (X. Wang and AHP, in preparation), with only ~3% differential sorghum–rice gene loss and less sorghum–rice structural rearrangement in 50 my than sorghum–maize rearrangement in ~10–15 my (Bowers et al. 2003). As has happened in maize, genome duplications and associated "fractionations" (gene losses) in other Andropogoneae crops have led to new gene linkage arrangements, and very probably to many new or modified gene functions (a hypothesis that we will test). In contrast, with 70 million years of "abstinence" from genome duplication, functions of *S. bicolor* genes may still

resemble those of the common cereal ancestor and therefore may be the best available "outgroup" for deducing the roles and evolutionary histories of duplicated, sub- or neo-functionalized genes in polyploid *S. halepense* as well as in maize, Saccharum, Miscanthus, and perhaps other major crops. Accordingly, the sequenced ~740 Mb sorghum genome (Paterson et al. 2009) is a logical complement to that of *Oryza* (rice) for grass functional genomics, as a representative of tropical grasses in which biochemical and morphological specializations ("C4" photosynthesis; see Wang et al., Chap. 19) improve carbon assimilation at high temperatures.

Particularly interesting may be the further investigation of karyotype evolution in the Saccharinae and their near relatives. The Saccharinae group of grasses shows an apparent reduction in chromosome number from the ancestral 20 to 10 in most parasorghums (Spangler et al. 1999) with another such reduction from 20 to 10 also experienced in parallel by the maize genome following whole-genome duplication since its divergence from *Sorghum* (Swigonova et al. 2004b); at least one and perhaps two chromosome doublings in *Saccharum* since its divergence from the remainder of the group (Ming et al. 1998); and both natural [*Sorghum halepense*: (Paterson et al. 1995b)] and human-mediated polyploidization [*Saccharum* cultivars (Ming et al. 1998)].

Knowledge of the mechanisms, levels, and patterns of evolution of genome size and structure in this curious group will be of especially great importance in laying the foundation for further study of the complex genomes of sugarcane and Miscanthus. As detailed in Chap. 1, Miscanthus and Saccharum are more closely related to one another than either is to sorghum. As detailed in Chap. 18, each has experienced genome duplication since divergence from a common ancestor shared with sorghum; however *Miscanthus* species have a basal set of 19 chromosomes ($2n = 38$, and 38 or 76 for spp. *sinensis* and *sacchariflorus*, respectively), versus the 10 that is characteristic of many Saccharinae including *Saccharum*. A fascinating question is whether genome duplication occurred once in a common ancestor of Miscanthus and Saccharum, or twice independently after their divergence. Their basal chromosome number of 19 suggests that *Miscanthus* homologs may have diverged sufficiently that they no longer normally pair with one another (perhaps including a chromosomal fusion to get from 20 to 19), unlike those of *Saccharum* that is largely autopolyploid. If indeed Miscanthus and Saccharum shared a genome duplication event, then their current karyotypes and pairing behavior imply that the two species have arrived at different "outcomes" from the same genome duplication, a fascinating possibility for which this author knows of no precedent.

3 Agriculture

The past few years have seen a reawakening of the developed world to the importance of agriculture, with a transition from chronic crop surpluses to very thin supplies, sharp increases in prices of many commodities, and both market-driven and

policy-driven incentives to increase the role of agriculture as a supplier of fuel. The sharp twenty-first century rise in food and energy prices may be a harbinger of the need anticipated by many leading agricultural scientists for a new Green Revolution (Conway 1997; Cantrell and Hettel 2004; Tuberosa et al. 2005). However, poor performance of Green Revolution rice cultivars in regions dependent upon low-input production necessitates that a "second" Green Revolution be particularly aimed at rainfed areas of Asia and Africa that are also contributing to ongoing population growth. More generally, increasing the harvest of solar energy through agriculture, for food, fuel, and other purposes, must also consider chronic problems such as soil erosion and runoff closely associated with intensive row-crop agriculture (Pimentel et al. 1995), as well as water deficits already limiting food production in many developing regions (Serageldin 2004), and predicted to challenge the food supply for 40% of the world's projected population by 2050 (UNESCO 2002). I believe that the Saccharinae are particularly well positioned to make greater than average contributions to these needs, as exemplified below, and therefore will grow in importance.

3.1 Drought

Sorghum is the most drought-tolerant dual-purpose (grain + straw) cereal crop of the semiarid tropics and subtropics, essential to human populations in arid or semiarid tropical regions such as the African Sahel and peninsular India. Despite its importance to resource-poor farmers living in drought-prone environments, sorghum improvement has lagged that of maize, wheat, and rice, each of which more than doubled in average yield on a worldwide basis from 1961 to 2001, during which sorghum yields only increased 51%. This gap needs to be remedied, through intensified research, deeper exploration of natural diversity, and prudent utilization of biotechnology. In particular, further enhancing its drought tolerance is a priority to improving food security in many regions inadequately addressed in the "first" Green Revolution, as water deficits remain the major factor contributing to instability of sorghum production globally.

QTL mapping and other gene tagging procedures have identified flanking markers for several sorghum target traits, including tolerance to pre-flowering drought stress (Tuinstra et al. 1996) and the stay-green component of terminal drought tolerance (Tuinstra et al. 1997; Tuinstra et al. 1998; Crasta et al. 1999; Subudhi and Nguyen 2000; Subudhi et al. 2000; Xu et al. 2000; Kebede et al. 2001; Haussmann et al. 2002; Sanchez et al. 2002; Harris et al. 2007). The "stay-green" trait is the best-characterized component of post-flowering drought tolerance in sorghum, contributing to grain and stover yield maintenance under terminal drought stress, and improving livestock feed value of straw (Hash et al. 2003).

The timing of flowering is of central importance to drought avoidance, adapting many grasses to seasonal fluctuation in rainfall. Diverse cereals share maximal sensitivity to water deficit at flowering. Most cereals (both wild and cultivated) originating

in the semiarid tropics are photoperiod sensitive and flower under short days (less than 12 h), coordinating seed development with favorable rainfall, temperature, and solar radiation conditions (Harper 1977). To adapt these and other cereals to temperate agriculture, photoperiod-insensitive (day-neutral) mutations have been selected that flower based on accumulation of "heat units" rather than under a specific day length. The need for different flowering regimes is a hindrance both in using exotic germplasm in temperate zones and in using improved temperate germplasm in the tropics. To mitigate these constraints for temperate regions, the "Sorghum Conversion Project" (Stephens et al. 1967; Rosenow and Clark 1987; Rosenow et al. 1997a, b; Dahlberg et al. 1998) has converted more than 673 exotic sorghums to day-neutral flowering by backcross introgression of photoperiod-insensitive mutations from a common donor (Tx406). The "converted" genotypes provide temperate-zone breeders with "user-friendly" access to exotic germplasm.

3.2 Perenniality

While annual crops have dominated the agricultural landscape since the time of the earliest farmers, benefits of perenniality are increasingly recognized. Perennial plants have four advantages over annuals: a longer growing season; better access to water and nutrients; more conservative use of nutrients; and better adaptation to marginal lands (Cox et al. 2006). The first three advantages are salient to all landscapes, from prime farmland to the poor or sloping soils that may be the only asset of a smallholder. By some estimates, one-third of the planet's arable land has been lost to soil erosion in the last few decades (Pimentel et al. 1995).

Parts of Africa including the Sahel and southern tip rank among the regions with most severe soil degradation worldwide (ISRIC 1990). Perennial cover is >50 times more effective than annual crops in maintaining topsoil (Gantzer et al. 1990). Consideration of how to expand agriculture to provide plant biomass for production of fuels or chemical feedstocks (Tilman et al. 2009) highlights many issues that are also high priorities to sustainable food production, particularly on depleted soils with a lack of available or affordable inputs such as irrigation, fertilizer, and hybrid seed (Glover et al. 2010).

Already a staple crop in regions of Africa with the most severe soil degradation, the importance of sorghum might be made even greater by the development of novel genotypes that increase the extent and duration of soil cover, mitigating disadvantages of conventional annual crops. Some sorghum cultivars can regrow from basal nodes after an initial harvest to produce a second "ratoon" crop, offering a potential production system that increases soil cover while reducing seed costs and energy expenditures (planting, tillage). Most such genotypes have ratoon crop yields of only ~50% of the seed crop (Duncan and Moss 1987); however there has been little if any research into whether this might be improved. Moreover, sorghum relatives *S. propinquum* and *S. halepense* are true perennials, the latter being able to overwinter as far north as Kansas, USA, by rhizomes, subterranean stems that serve both as

a carbohydrate repository and a vegetative propagule. Sorghum has become a botanical model for dissecting the molecular basis of perenniality (Paterson et al. 1995a; Hu et al. 2003; Jang et al. 2006; Jang et al. 2008), and it may be uniquely well positioned to become an early adopter of the potential benefits of ratoon/perennial crop production systems.

4 Weediness and Invasion Biology

Some perennials maximize the advantages detailed above (Cox et al. 2006) of this life history strategy to outcompete native vegetation on unmanaged and marginal lands and are deemed "invasive" (Rout and Chrzanowski 2009). With the tremendous mobility of humanity, plant invasions are a frequent and costly occurrence, yet remain only poorly understood. Just as it has become a botanical model for dissecting the molecular basis of perenniality, the Sorghum genus also may offer insights into invasion biology.

Sorghum halepense (L.) Pers. is a polyploid resulting from natural hybridization between *S. bicolor* ($2n=20$), which is an annual native to Africa, and *S. propinquum* (Kunth) Hitchc. ($2n=20$), a wild perennial from Southeast Asia estimated to have diverged from *S. bicolor* ~1–2 mya (Feltus et al. 2004). *S. halepense* finds appreciable use as a forage crop in many countries including the USA, and occasional use for food (seed/flour) in low-income countries. However, it is of greatest importance as one of the world's most noxious weeds, having spread from its west Asian center of diversity across much of Asia, Africa, Europe, North and South America, and Australia in a range from about 50°N to 45°S latitude that far exceeds the range to which its progenitor species are adapted, either in the wild or in cultivation. Its establishment in the USA is probably typical of its spread to other continents, including intentional introduction as a prospective forage and/or unintentional introduction as a contaminant of sorghum seedlots (McWhorter 1971). However, while *S. bicolor* has remained confined to cultivation, *S. halepense* has readily naturalized—suggesting genetic potential for adaptation that goes well beyond that of sorghum.

In addition to having spread more widely, *S. halepense* is also more polymorphic than its progenitors. In an exploratory genetic analysis, 18 *S. halepense/S. almum* genotypes from the USA (8), Chile, Argentina, Algeria, New Zealand, Australia (2), South Africa, India, Kazakhstan, and former USSR averaged 6.13 alleles per locus versus 3.39 for a worldwide sample of 55 landrace and wild sorghum accessions and 1.9 for 16 F1 hybrid sorghums from 8 US commercial breeding programs (Morrell et al. 2005). Within the USA, geographically distant populations are well differentiated in allele and haplotype frequencies (U. Sezen, H. Tang, AHP, unpubl.).

Early evidence suggests that there have also been changes in gene function in *S. halepense* relative to its progenitors. Genes that were shown to have rhizome-enriched expression in polyploid *S. halepense* (Jang et al. 2006) appeared to be more frequently derived from the *S. propinquum* than the *S. bicolor* progenitor, but there was evidence of formation of novel alleles and "recruitment" of *S. bicolor* genes to rhizome-enriched expression in *S. halepense* (Jang et al. 2008).

In partial summary, a hypothesis that the author is further exploring is that the post-polyploidy evolution of duplicated genes in *S. halepense* may have led to alleles conferring modified or even novel functions relative to those of their progenitor genes in *S. bicolor* and *S. propinquum,* and predisposing *S. halepense* to adapt to a host of environments that its diploid progenitors never experienced (or never adapted to).

5 Summation and Closing

It is my sincere hope that this volume has conveyed to you, the reader, some of the enthusiasm with which many colleagues (both coauthors and many others) and I pursue our work on the Saccharinae group and the many contributions that we feel this group is likely to make to benefit humanity. We hope that this volume provides a resource to quickly update others on the current status of Saccharinae research and resources, while also illustrating some past successes and future opportunities. Inevitably, any book reflects a subset of viewpoints in a community (naturally most heavily biased by my own, and to a lesser degree those of the chapter authors and coauthors), and we naturally recognize that there have been omissions herein. Indeed, limits on the time of busy professionals necessitated withdrawal of a few intended chapters. Nonetheless, we believe that this work provides a review and synthesis of Saccharinae research that will be of value to you, and hope in particular that these pages help to stimulate intensified research into and improvement of Saccharinae crops.

References

ISRIC (1990) The Global Assessment of Human Induced Soil Degradation (GLASOD) Digital Database from UNEP/GRID-Geneva. Wageningen, The Netherlands

Bowers JE, Abbey C, Anderson S, Chang C, Draye X, Hoppe AH, Jessup R, Lemke C, Lennington J, Li Z, Lin YR, Liu SC, Luo L, Marler BS, Ming R, Mitchell SE, Kresovich S, Schertz KF, Paterson AH (2003) A high-density genetic recombination map of sequence-tagged sites for sorghum, as a framework for comparative structural and evolutionary genomics of tropical grains and grasses. Genetics 165:367–386

Cantrell RP, Hettel GP (2004) Rice, biofortification, and enhanced nutrition. World Food Prize Symposium, Des Moines, IA

Conway G (1997) The doubly green revolution: food for all in the twenty-first century, Cornell University Press, Ithaca NY

Cox TS , Glover JD, Van Tassel DL, Cox CM, DeHaan LR (2006) Prospects for developing perennial gain crops. BioScience 56:649–659

Crasta OR, Xu WW, Rosenow DT, Mullet J, Nguyen HT (1999) Mapping of post-flowering drought resistance traits in grain sorghum: association between QTLs influencing premature senescence and maturity. Mol Gen Genet 262:579–588

Dahlberg J, Rosenow DT, Peterson GC, Clark LE, Miller FR, Sotomayor RA, Hamburger AJ, Madera Torres P, Quiles Belen A, Woodfin CA (1998) Registration of 40 converted sorghum germplasms. Crop Sci 38:564–565

Duncan RR, Moss RB (1987) Comparative yields of ratoon cropped temperately and tropically adapted grain-sorghum hybrids. Crop Sci 27:569–571

Feltus FA, Wan J, Schulze SR, Estill JC, Jiang N, Paterson AH (2004) An SNP resource for rice genetics and breeding based on subspecies Indica and Japonica genome alignments. Genome Res 14:1812–1819

Gantzer CJ, Anderson SH, Thompson AL, Brown JR (1990) Estimating soil erosion after 100 years of cropping on Sanborn Field. J Soil Water Conserv 45:641–644

Gaut BS, Clark LG, Wendel JF, Muse SV (1997) Comparisons of the molecular evolutionary process at rbcL and ndhF in the grass family (Poaceae). Mol Biol Evol 14:769–777

Glover JD, Reganold JP, Bell LW, Borevitz J, Brummer EC, Buckler ES, Cox CM, Cox TS, Crews TE, Culman SW, DeHaan LR, Eriksson D, Gill BS, Holland J, Hu F, Hulke BS, Ibrahim AMH, Jackson W, Jones SS, Murray SC, Paterson AH, Ploschuk E, Sacks EJ, Snapp S, Tao D, Van Tassel DL, Wade LJ, Wyse DL, Xu Y (2010) Increased food and ecosystem security via perennial grains. Science 328:1638–1639

Harper J (1977) Plant population biology. Academic, London

Harris K, Subudhi PK, Borrell A, Jordan A, Rosenow DT, Nguyen HT, Klein PE, Klein RR, Mullet J (2007) Sorghum stay-green QTL individually reduce post-flowering drought-induced leaf senescence. J Exp Bot 58:327–338

Hash CT, Raj AGB, Lindup S, Sharma A, Beniwal CR, Folkertsma RT, Mahalakshmi V, Zerbini E, Blummel M (2003) Opportunities for marker-assisted selection (MAS) to improve the feed quality of crop residues in pearl millet and sorghum. Field Crop Res 84:79–88

Haussmann BIG, Mahalakshmi V, Reddy BVS, Seetharama N, Hash CT, Geiger HH (2002) QTL mapping of stay-green in two sorghum recombinant inbred populations. Theor Appl Genet 106:133–142

Heaton EA, Dohleman FG, Long SP (2008) Meeting US biofuel goals with less land: the potential of Miscanthus. Glob Chang Biol 14:2000–2014

Hu FY, Tao DY, Sacks E, Fu BY, Xu P, Li J, Yang Y, McNally K, Khush GS, Paterson AH, Li ZK (2003) Convergent evolution of perenniality in rice and sorghum. Proc Natl Acad Sci U S A 100:4050–4054

Jang CS, Kamps TL, Skinner DN, Schulze SR, Vencill W, Paterson AH (2006) Sorghum genes with rhizome-enriched expression: functional classification, genomic organization, putative cis-acting regulatory elements, and relationship to QTLs. Plant Physiol 142:1148–1159

Jang CS, Kamps TL, Tang H, Bowers JE, Lemke C, Paterson AH (2008) Evolutionary fate of rhizome-specific genes in a non-rhizomatous Sorghum genotype. Heredity 102:266–273

Jannoo N, Grivet L, Chantret N, Garsmeur O, Glaszmann JC, Arruda P, D'Hont A (2007) Orthologous comparison in a gene-rich region among grasses reveals stability in the sugarcane polyploid genome. Plant J 50:574–585

Kebede H, Subudhi PK, Rosenow DT, Nguyen HT (2001) Quantitative trait loci influencing drought tolerance in grain sorghum (Sorghum bicolor L. Moench). Theor Appl Genet 103:266–276

McWhorter CG (1971) Introduction and spread of Johnsongrass in the United States. Weed Sci 19:496

Ming R, Liu SC, Lin YR, da Silva J, Wilson W, Braga D, van Deynze A, Wenslaff TF, Wu KK, Moore PH, Burnquist W, Sorrells ME, Irvine JE, Paterson AH (1998) Detailed alignment of Saccharum and Sorghum chromosomes: comparative organization of closely related diploid and polyploid genomes. Genetics 150:1663–1682

Morrell PL, Williams-Coplin D, Bowers JE, Chandler JM, Paterson AH (2005) Crop-to-weed introgression has impacted allelic composition of johnsongrass populations with and without recent exposure to cultivated sorghum. Mol Ecol 14:2143–2154

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Lyons E, Maher C, Narechania A, Penning B, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein PE, Kresovich S, McCann MC, Ming R, Peterson DG, Ware D, Westhoff P, Mayer KFX, Messing J, Rokhsar DS (2009) The Sorghum bicolor genome and the diversification of grasses. Nature 457:551–556

Paterson AH, Schertz KF, Lin YR, Liu SC, Chang YL (1995a) The weediness of wild plants—molecular analysis of genes influencing dispersal and persistence of Johnsongrass, Sorghum halepense (L) Pers. Proc Natl Acad Sci U S A 92:6127–6131

Paterson AH, Schertz KF, Lin YR, Liu SC, Chang YL (1995b) The weediness of wild plants—molecular analysis of genes influencing dispersal and persistence of johnsongrass, Sorghum-halepense (L) Pers. Proc Natl Acad Sci U S A 92:6127–6131

Pimentel D, Harvey C, Resosudarmo P, Sinclair K, Kurz D, McNair M, Crist S, Shpritz L, Fitton L, Saffouri R, Blair R (1995) Environmental and economic costs of soil erosion and conservation benefits. Science 267:1117–1123

Rosenow DT, Clark LE (1987) Utilization of exotic germplasm in breeding for yield stability. Fifteenth biennial grain sorghum research and utilization conference, pp 49–56

Rosenow DT, Dahlberg J, Stephens JC, Miller FR, Barnes DK, Peterson GC, Johnson JW, Schertz KF (1997a) Registration of 63 converted sorghum germplasm lines from the sorghum conversion program. Crop Sci 37:1399–1400

Rosenow DT, Dahlberg JA, Peterson GC, Clark LE, Miller FR, Sotomayor RA, Hamburger AJ, Madera Torres P, Quiles Belen A, Woodfin CA (1997b) Registration of fifty converted sorghums from the sorghum conversion program. Crop Sci 37:1397–1398

Rout M, Chrzanowski T (2009) The invasive Sorghum halepense harbors endophytic N2-fixing bacteria and alters soil biogeochemistry. Plant Soil 315:163–172

Sanchez AC, Subudhi PK, Rosenow DT, Nguyen HT (2002) Mapping QTLs associated with drought resistance in sorghum (Sorghum bicolor L. Moench). Plant Mol Biol 48:713–726

Serageldin I (2004) Speculations on the future of water and food security. International Food Policy Research Institute, Washington, DC

Sobral BWS, Braga DPV, Lahood ES, Keim P (1994) Phylogenetic analysis of chloroplast restriction enzyme site mutations in the Saccharinae Griseb subtribe of the Andropogoneae Dumort tribe. Theor Appl Genet 87:843–853

Spangler R, Zaitchik B, Russo E, Kellogg E (1999) Andropogoneae evolution and generic limits in Sorghum (Poaceae) using ndhF sequences. Syst Bot 24:267–281

Stephens J, Miller F, Rosenow D (1967) Conversion of alien sorghums to early combine genotypes. Crop Sci 7:396

Subudhi PK, Nguyen HT (2000) Linkage group alignment of sorghum RFLP maps using a RIL mapping population. Genome 43:240–249

Subudhi PK, Rosenow DT, Nguyen HT (2000) Quantitative trait loci for the stay green trait in sorghum (Sorghum bicolor L. Moench): consistency across genetic backgrounds and environments. Theor Appl Genet 101:733–741

Swigonova Z, Lai J, Ma J, Ramakrishna W, Llaca V, Bennetzen JL, Messing J (2004a) Close split of sorghum and maize genome progenitors. Genome Res 14:1916–1923

Swigonova Z, Lai JS, Ma JX, Ramakrishna W, Llaca M, Bennetzen JL, Messing J (2004b) On the tetraploid origin of the maize genome. Comp Funct Genomics 5:281–284

Tilman D, Socolow R, Foley JA, Hill J, Larson E, Lynd L, Pacala S, Reilly J, Searchinger T, Somerville C, Williams R (2009) Beneficial biofuels—the food, energy, and environment trilemma. Science 325:270–271

Tuberosa R, Phillips RL, Gale MD (2005) In the wake of the double helix: from the green revolution to the gene revolution. Avenue Media, Bologna

Tuinstra MR, Ejeta G, Goldsbrough P (1998) Evaluation of near-isogenic sorghum lines contrasting for QTL markers associated with drought tolerance. Crop Sci 38:835–842

Tuinstra MR, Grote EM, Goldsbrough PB, Ejeta G (1996) Identification of quantitative trait loci associated with pre-flowering drought tolerance in sorghum. Crop Sci 36:1337–1344

Tuinstra MR, Grote EM, Goldsbrough PB, Ejeta G (1997) Genetic analysis of post-flowering drought tolerance and components of grain development in Sorghum bicolor (L.) Moench. Mol Breed 3:439–448

UNESCO (2002) Vital water graphics, water use and management. United Nations Education Scientific and Cultural Organization, Paris

Xu WW, Subudhi PK, Crasta OR, Rosenow DT, Mullet JE, Nguyen HT (2000) Molecular mapping of QTLs conferring stay-green in grain sorghum (Sorghum bicolor L. Moench). Genome 43: 461–469

Index

A.H. Paterson (ed.), *Genomics of the Saccharinae*, Plant Genetics and Genomics: Crops and Models 11, DOI 10.1007/978-1-4419-5947-8,

N

O

P

Q