Xinyou Yin · Paul C. Struik *Editors*

Crop Systems Biology

Narrowing the Gaps between Crop Modelling and Genetics

Springer

Crop Systems Biology

Xinyou Yin • Paul C. Struik
Editors

Crop Systems Biology

Narrowing the Gaps between Crop Modelling and Genetics

Springer

Editors
Xinyou Yin
Department of Plant Sciences,
Centre for Crop Systems Analysis
Wageningen University
Wageningen, The Netherlands

Paul C. Struik
Department of Plant Sciences,
Centre for Crop Systems Analysis
Wageningen University
Wageningen, The Netherlands

ISBN 978-3-319-20561-8 ISBN 978-3-319-20562-5 (eBook)
DOI 10.1007/978-3-319-20562-5

Library of Congress Control Number: 2015953437

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Preface

The sequencing of genomes has been completed for an increasing number of crop species, and researchers have now succeeded in isolating and characterising many important genes and quantitative trait loci. High expectations from genomics, however, are waving back towards the recognition that crop physiology is also important for realistic improvement of crop productivity. Complex processes and networks along various hierarchical levels of crop growth can be thoroughly understood with the help of their mathematical description – modelling. The further practical application of these understandings also requires the discovery of emerging properties and quantitative predictions. In order to better support design, engineering and breeding for new crops and cultivars for improving agricultural production under global warming and climate change, there is an increasing call for an interdisciplinary research approach, which combines modern genetics and genomics, traditional physiology and biochemistry and advanced bioinformatics and crop modelling. Recently we coined a term 'crop systems biology' to describe such an interdisciplinary approach.

Such an interdisciplinary research has been practised in various research groups across the globe. However, it does not seem to be fully covered in the format of book publications. We, therefore, initiated this book project on 'crop systems biology – narrowing the gaps between crop modelling and genetics', in response to an invitation by Springer Science + Business Media. Nine chapters written by leading groups active in this field are presented in the book, representing the state of the art in the realm of this research covering various traits in several crops.

Baldazzi et al. describe gene regulatory and metabolic networks, link these networks to crop models and show how to integrate different temporal and spatial scales within a single model, thus illustrating the perspectives for multi-scale modelling.

Xu and Buck-Sorlin describe a three-dimensional modelling approach called Functional-Structural Plant Modelling and link it to quantitative trait loci analysis in rice. They use this framework to analyse opportunities and pitfalls to advance breeding for architectural traits.

Bustos et al. discuss prediction strategies for genotype-by-environment interactions using statistical models, crop growth models and combinations of

model types. They illustrate how prediction accuracy can profit from the large data sets available on environmental and genotypic variables by integrating physiological and statistical knowledge.

Génard et al. show how knowledge generated by in silico profiling can be used to unravel genotype × environment × management interactions and to construct plant ideotypes for particular conditions, using examples for fruit quality, sensitivity to diseases and root system architecture.

Luquet et al. describe a model that combines characteristics of functional-structural modelling approaches with classical crop growth models. They use this model to analyse the trade-off between early vigour and drought tolerance in rice and to design rice ideotypes that combine the two traits.

Sinclair et al. describe the steps of modelling-physiology-transcriptomics-genetic screening they followed in developing soybean cultivars with restricted transpiration. The yield increases obtained in experiments and in model simulations for years with limited rainfall prove that this trait is highly desirable.

Hammer et al. argue that crop ecophysiology and functional modelling can effectively link processes at the molecular and organism levels. They provide a physiological framework and examples, illustrating that their integrated functional modelling and molecular genetics approach holds promise for closing the genotype-to-phenotype gap.

Boote et al. show opportunities and challenges of linking genetics to process-oriented crop modelling, with the objective of predicting field performance of grain legumes as a function of genes. They also show how to link model input parameters with allelic effects of several known genes to predict growth and seed yield in the common bean.

Yin et al. describe the most active research line within crop systems biology over the last 15 years: quantitative trait loci-based crop modelling; they provide a comprehensive overview of recent experiences and future prospects within this field.

In the last chapter, the editors outline how these research activities contribute to the development of crop systems biology within the context of crop improvement programmes.

The book is meant for those scientists and graduate students from the domains of, and interested in bridging, fundamental plant biology and applied crop science.

As presented in the book, crop systems biology is a dynamically evolving concept and research realm. We appreciate receiving any response and feedback from readers. Please do not hesitate to contact us if you have suggestions or comments.

We thank the authors of individual chapters for their valuable contribution. We also thank Maryse Elliott, Melanie van Overbeek and Anja Smykowski of Springer Science + Business Media B.V. for inviting us to initiate this project and for their subsequent support in realising it.

Wageningen, The Netherlands — Xinyou Yin
March 2015 — Paul C. Struik

Contents

Contributors

Valentina Baldazzi INRA, UR1115 Plantes et Systèmes de Culture Horticoles, Avignon, France

Nadia Bertin INRA, UR1115 Plantes et Systèmes de Culture Horticoles, Avignon, France

Kenneth J. Boote Agronomy Department, University of Florida, Gainesville, FL, USA

Andrew Borrell Centre for Plant Science, Queensland Alliance for Agriculture and Food Innovation, Hermitage Research Facility, The University of Queensland, Warwick, Australia

Jacques Le Bot INRA, UR1115 Plantes et Systèmes de Culture Horticoles, Avignon, France

Gerhard Buck-Sorlin UMR 1345 Institut de Recherche en Horticulture et Semences (IRHS), AGROCAMPUS OUEST, Angers Cedex 01, France

Daniela Bustos-Korts Biometris, Wageningen University & Research Centre, Wageningen, The Netherlands

C.T. de Wit Graduate School for Production Ecology and Resource Conservation (PE&RC), Wageningen, The Netherlands

Thomas E. Carter Jr. Agricultural Research Service, US Department of Agriculture, Raleigh, NC, USA

Scott Chapman CSIRO Plant Industry and Climate Adaptation Flagship, Queensland Bioscience Precinct, St Lucia, QLD, Australia

Mark Cooper DuPont-Pioneer, Johnston, IA, USA

Melanie J. Correll Agricultural and Biological Engineering Department, University of Florida, Gainesville, FL, USA

Elsa Desnoues INRA, UR1115 Plantes et Systèmes de Culture Horticoles, Avignon, France

Jyostna M. Devi Crop Science Department, North Carolina State University, Raleigh, NC, USA

Michael Dingkuhn UMR AGAP, CIRAD, Montpellier, France

CESD Division, IRRI, Metro Manila, Philippines

Fred van Eeuwijk Biometris, Wageningen University & Research Centre, Wageningen, The Netherlands

Hélène Gautier INRA, UR1115 Plantes et Systèmes de Culture Horticoles, Avignon, France

Michel Génard INRA, UR1115 Plantes et Systèmes de Culture Horticoles, Avignon, France

Junfei Gu Key Laboratory of Crop Genetics and Physiology of Jiangsu Province, Yangzhou University, Yangzhou, Jiangsu, China

Graeme Hammer Centre for Plant Science, Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Brisbane, Australia

James W. Jones Agricultural and Biological Engineering Department, University of Florida, Gainesville, FL, USA

David Jordan Centre for Plant Science, Queensland Alliance for Agriculture and Food Innovation, Hermitage Research Facility, The University of Queensland, Warwick, Australia

Françoise Lescourret INRA, UR1115 Plantes et Systèmes de Culture Horticoles, Avignon, France

Delphine Luquet UMR AGAP, CIRAD, Montpellier, France

Marcos Malosetti Biometris, Wageningen University & Research Centre, Wageningen, The Netherlands

Mohamed-Mahmoud Memmah INRA, UR1115 Plantes et Systèmes de Culture Horticoles, Avignon, France

Charlie Messina DuPont-Pioneer, Johnston, IA, USA

Erik van Oosterom Centre for Plant Science, Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Brisbane, Australia

Loïc Pagès INRA, UR1115 Plantes et Systèmes de Culture Horticoles, Avignon, France

Bénédicte Quilot-Turion INRA, UR1052 Génétique et Amélioration des Fruit et Légumes, Avignon, France

Camila Rebolledo CIAT, Agrobiodiversity, Cali, Colombia

Lauriane Rouan UMR AGAP, CIRAD, Montpellier, France

Thomas R. Sinclair Crop Science Department, North Carolina State University, Raleigh, NC, USA

Vijaya Singh Centre for Plant Science, Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Brisbane, Australia

Jean-Christophe Soulie UMR AGAP, CIRAD, Montpellier, France

Paul C. Struik Centre for Crop Systems Analysis, Department of Plant Sciences, Wageningen University, Wageningen, The Netherlands

C. Eduardo Vallejos Horticultural Sciences Department, University of Florida, Gainesville, FL, USA

Gilles Vercambre INRA, UR1115 Plantes et Systèmes de Culture Horticoles, Avignon, France

Huaqi Wang Plant Breeding and Genetics, China Agricultural University, Beijing, People's Republic of China

Lifeng Xu College of Computer Science & Technology, Zhejiang University of Technology, Hangzhou, People's Republic of China

Xinyou Yin Department of Plant Sciences, Centre for Crop Systems Analysis, Wageningen University, Wageningen, The Netherlands

Chapter 1
Challenges in Integrating Genetic Control in Plant and Crop Models

Valentina Baldazzi, Nadia Bertin, Michel Génard, Hélène Gautier, Elsa Desnoues, and Bénédicte Quilot-Turion

Abstract Predicting genotype-to-phenotype relationships under contrasting environments is a great challenge for plant biology and breeding. Classical crop models have been developed to predict crop yield or product quality under fluctuating environments but they are usually calibrated for a single genotype, restricting the validity range of the model itself. To overcome this limitation, genetic control has to be integrated into crop models and genotype×environment interactions have to be made explicit.

The aim of this chapter is to provide an overview of a panel of approaches that have been developed to integrate genetic control in plant and crop models. The fundamentals of quantitative genetics of complex traits are first introduced, with special attention to methods for quantitative trait loci (QTL) cartography and QTL genetic modelling. The integration of genetic control within ecophysiological models is then discussed. Several methods are reviewed, ranging from classical statistical approaches, relying on specific model parameters reflecting gene or QTL effects, to recent multi-scale models, explicitly integrating molecular networks. This chapter proposes a review of a few techniques from systems biology that can be used to describe the behaviour of cellular networks, in a simplified way. Coupling different organizational scales is finally discussed and a few examples of multi-scale plant models are presented.

V. Baldazzi (✉) • N. Bertin • M. Génard • H. Gautier • E. Desnoues
INRA, UR1115 Plantes et Systèmes de Culture Horticoles, 84000 Avignon, France
e-mail: valentina.baldazzi@avignon.inra.fr

B. Quilot-Turion
INRA, UR1052 Génétique et Amélioration des Fruit et Légumes, 84000 Avignon, France

X. Yin, P.C. Struik (eds.), *Crop Systems Biology*,
DOI 10.1007/978-3-319-20562-5_1

1.1 Introduction

Predicting the behaviour of a given genotype under contrasting environments or predicting the phenotypes of contrasting genotypes under one environment is difficult for various reasons. First, small genome variations may have huge impacts on the phenotype due to cascade effects on the processes involved in the observed traits. Second, strong genotype-by-environment (G×E) interactions occur for most plant traits. During the last decades, many process-based models have been developed to predict crop yield or quality under fluctuating environments (Martre et al. 2011). These models usually describe the temporal variations of main processes involved in final traits as well as their interactions and responses to environmental variations or cultural practices. Yet, most of the processes involved in crop yield or quality depend also on the genetic makeup with high environment-by-genotype interactions (e.g., Kromdijk et al. 2014; Prudent et al. 2010). This implies that model parameters are usually specific for one genotype, restricting the validity range of the model itself. To overcome this limitation, several authors tried to take into account the genetic control in plant models. This implies to identify the model parameters which are genotype dependent and to quantify them depending on combinations of Quantitative Trait Loci (QTL) (QTL-based models), alleles or genes (gene-based models) involved in the process which is modelled. The complexity of the genetic control accounted for in models is usually the inverse of the complexity of the modelled system. Indeed integrating the effects of many genes is possible at the cell or organ level, more rarely at the plant level. However, methods developed at the lower levels of organization may teach us how to proceed through higher levels. Plant complexity cannot be described in one big multi-scale model. Instead we should now focus on (i) tracing main hubs at the lower levels of organization; (ii) quantifying their effects at the higher levels of organization; (iii) refining the plant models to be able to link model parameters to physiological components. This will be a real challenge for the future due to the polygenic control of most of the variables and processes involved in crop yield and product quality and to the strong genotype-by-environment interactions. The success will rely on the advance in the understanding of the genetic control of the target traits and on our ability to phenotype large populations under contrasting environments at the process level.

The overall objective of this chapter is to provide an overview of the integration of genetic control in plant and crop models at different levels of organization, from gene networks, to cell, organ and plant.

The chapter is organized as follows: the fundamentals of quantitative genetics of complex traits are first introduced, with special attention to methods for QTL cartography and QTL genetic modelling. The integration of genetic control within ecophysiological models is then discussed. Classical approaches are first introduced that rely on the specification of model parameters as a function of QTL or gene effects using simple empirical relations. The advantage of this approach resides in its flexibility and in the possibility to describe experimental data, even in absence of a clear understanding of the underlying biological mechanisms.

The last part of the chapter is devoted to emerging multi-scale models, explicitly integrating the description of molecular processes with a broader view to plant physiology and development. In this perspective, Sect. 1.3 proposes a review of a few selected methods from systems biology that can be used to describe the behaviour of cellular network. Model simplification and coupling among different organizational scales is the topic of Sect. 1.4.

1.2 Fundamentals in Quantitative Genetics

Here is a list of definitions that may help the reader to follow the subsequent paragraphs.

Allele	Functional form of a gene
Locus	Particular site of a gene or DNA sequence on a chromosome
Homozygosity	In case of diploid plants, both alleles have the same phenotypic effect
Heterozygosity	Each allele at a gene has a different phenotypic effect
Additive allele effect (a)	In a population, a is half of the difference between the trait value (Y) of the mean homozygous genotypes for one parental allele (mm) and for the other parental allele (MM): $a=(Y_{mm}-Y_{MM})/2$
Dominance effect	Accounts for the interaction between alleles at each locus. This term is non null when heterozygotes (a A) are not exact midpoints of homozygote parents. It is the deviation between the mean of heterozygotes and the half sum of the homozygotes
Epistasis	The alleles at one locus change the phenotypic effects of genetic variation at another locus
Plasticity	A single genotype gives rise to a diversity of phenotypes, depending on specific environmental conditions or lifetime
Polygeny	Many genes contribute to a particular phenotypic character
Pleiotropy	Different phenotypic characters are affected by a single genetic variation
Transgressive individuals	An individual has a phenotype more extreme than the phenotype displayed by the two parents

1.2.1 Quantitative Traits Controlled by Numerous Genes

One of the fundamental ideas of quantitative genetics, as defined by Fisher in 1918, is that the phenotypic value P of an individual is the sum of that individual's genotypic value G plus its environmental value E: $P=G+E$. The genotypic value is the combined effect of all the genetic effects, including nuclear and mitochondrial

genes, and interactions between the genes. The average phenotypic outcome may be affected by dominance and by how genes interact with genes at other loci.

In case of clones, the value G corresponds to the average value of all clones whereas the environmental value E is inferred from the difference between the phenotype of the clone and G. However, G and E can also be estimated in case of other types of related individuals. This is the common case where G values are different between individuals and depend on their relations. To account for this, the genotypic value is decomposed as:

$$G = A + D + I$$

where A represents the contribution to the character from the effects of individual alleles, D is the contribution from the interaction between these alleles, and I represents the contribution from interactions between different loci.

In summary, the genetic architecture of complex traits first implies the actions of genes in singular locus but also the inter-locus interactions and gene × environment interactions.

Quantitative genetics developed first from the analysis of characters with discrete variations. The determinism of such characters often proved to be monogenic characterized by the wild phenotype and the mutant. In most cases, those phenotypes were caused by a major alteration of a single gene. However those alterations are quite rare since they are counter-selected because of their large phenotypic effect.

Indeed, many characters show continuous variations in the populations. Those characters are under polygenic control: many loci called QTL (Quantitative Trait Loci) are responsible for the observed variations. Their nature may not be different from the one of the loci responsible for discrete variations. Indeed the main difference lies in the moderate effects of each locus. Most of the traits of interest are controlled by multiple interacting genes. So up to now the huge progress in gene discovery has only weakly aided genetic selection (Miflin 2000; Sinclair et al. 2004). For instance, in tomato (*Solanum lycopersicum* L.) fruit, more than 100 genes located in 16 regions of the genome, are associated with fruit composition, mainly sugar and acid contents (Causse et al. 2004; Bermudez et al. 2008). Consequently, QTL for a given trait usually explain only low proportions of the observed trait variations. Moreover, most of these QTL depend on the environment (QTL × E) and on the genetic background (QTL × QTL) (Börner et al. 1993; Blanco et al. 2002; Chaïb et al. 2006; Causse et al. 2007; Dudley et al. 2007).

Current technological progresses and recent advances in genetic analyses may offer possibilities to estimate more and more precisely the individual effects of each locus detected, their location on the genome and their potential interactions with other loci. This information is precious to build a model that predicts the value of an individual with a given combination of alleles.

1.2.2 Principles and Methods of QTL Cartography

Mapping QTL is based on the systematic search for associations between marker loci and the quantitative traits (Kearsey 1998). The prerequisites for QTL mapping are:

- create a segregating progeny. Best efficiency is reached in case of crosses between inbred lines.
- for each individual of the progeny, get the genotype of a set of marker loci distributed along the genome, so as to build a genetic map of the progeny.
- get the value of the studied trait for each individual of the progeny.
- perform biometric methods to detect an association between the score of marker genotypes and the value of the measured trait and estimate the genetic parameters of the detected QTL.

In the last decades, the tremendous advances in molecular genetics have greatly facilitated genetic analysis of quantitative traits. More recently, the use of markers based on single nucleotide polymorphisms (SNPs) have rapidly increased in plant genetics due to their abundance in the genomes and the possibility of high-throughput detection (Mammadov et al. 2012). It is now fairly routine to locate highly polymorphic marker loci that span the genome. Consequently, the major challenge is now the phenotypic analysis of the genetic variability (Houle et al. 2010), which requires simultaneous analyses of hundreds to thousands of plants. To face this difficulty, phenotypic platforms allowing fine environmental control (Tisné et al. 2013; Granier et al. 2006) or field characterization (Andrade-Sanchez et al. 2013) are emerging.

The simplest method to detect QTL is to consider the molecular markers independently. A difference in the mean trait value between different marker genotypes is sought. However, if the marker and QTL are separated by some recombination fractions, the strength of the marker–trait association decreases. Thus, a weak association can be generated by tight linkage to a QTL of small effect or loose linkage to a QTL of major effect. To further decipher these cases, different statistical approaches have been used that allow estimating QTL effects and their map positions.

The premise of identifying QTL is based on the likelihood ratio of the probability of having an association between a marker and a QTL assuming genetic linkage, divided by the probability of having an association assuming no linkage. This ratio is called LOD (logarithm of the odds) score. A LOD score of 3 or greater is usually considered as statistically significant evidence for linkage between a marker and a QTL. However, different methods are available to calculate a genome-wide significance threshold, from permutation tests to Bayesian approaches.

A number of statistical methods have been developed for mapping QTL, from marker by marker analysis (variance analysis, Student test) to multi-environment mapping. Based on maximum likelihood algorithms, Lander and Botstein (1989)

proposed what is now called interval mapping (IM) to scan the genome for evidence of QTL. IM can also be performed by regression (Knapp et al. 1990; Haley and Knott 1992). Subsequently, composite interval mapping (CIM) was developed. The method described by Zeng (1993) is based on multiple regressions to isolate individual QTL effects and genetic variation in other regions of the genome. The aim is to reduce the background "noise" that can affect QTL detection by incorporating into the model a set of markers significantly associated with the trait. These 'cofactors' may be located anywhere in the genome. Jansen and Stam (1994) also proposed 'multiple QTL model' (MQM), a method similar to CIM. Compared with IM, both CIM and MQM can significantly improve mapping precision and the estimation of QTL effects by the fact that more QTL are detected.

Advanced statistical methods, e.g., to perform multi-environment and/or multitrait QTL mapping, have emerged and have recently been reviewed by van Eeuwijk et al. (2010). They state that the mixed model QTL methodology is suitable for many types of populations and allows predictive modelling of QTL by environment interactions.

In parallel, other statistical methods have been developed to detect QTL involved in response curves ('functional mapping'). For example, Ma et al. (2002) combined logistic growth curves and QTL mapping within a mixture model approach. This method proved to be powerful and to produce accurate estimates of QTL effects and positions (Wu et al. 2002, 2003). Using a similar framework, Malosetti et al. (2006) proposed a non-linear extension of classical mixed models.

1.2.3 QTL Genetic Parameters

The biometric methods cited above not only detect QTL (map location) but also estimate a number of genetic parameters of these QTL. Among them, the QTL effect is the difference of effect between alleles, usually referred to as the additive effect (a) or effect of a double substitution (2a). In addition to the magnitude, the sign of the effect is also of particular interest. Indeed, both favorable and unfavorable alleles sometimes come from the two parents. In addition, the QTL effect can also refer to the part of the phenotypic variation explained by each QTL or by all QTL controlling a trait. This part (R^2) is quantified by the percentage of the difference between the residual sum of squares (RSS) of the reduced model and the full model, divided by the full model RSS.

Based on the QTL analysis results, a quantitative genetic model can be defined that relates the genotypic value of an individual to the alleles at the loci that contribute to the variation in a population in terms of additive, dominance, and epistatic effects. For example, Podlich and Cooper (1998) developed a platform for quantitative analysis of genetic models, QU-GENE. The definition of the genetic model includes the following components:

1. Number of genes (loci).
2. Intra-locus gene action (additive, dominance).

3. Inter-locus gene action (epistasis).
4. Pleiotropy.
5. Number of alleles.
6. Gene frequency (allele frequency).
7. Mutation.
8. Ploidy.
9. Linkage and chromosomal arrangements.
10. Genotype-by-environment interaction.

From this genetic model, the genotypic value of any individual genotype, carrying any combination of alleles from this population, may be inferred. Reymond et al. (2003) used a QTL model including additive and epistatic effects. Then they estimated for each individual the allelic probability at QTL positions, given the information at flanking markers, and finally used them in the QTL model.

1.3 Integration of Genetic Control in Crop Models

1.3.1 Levels of Integration

White and Hoogenboom (2003) reviewed the issues related to incorporating gene action into crop models. They proposed a classification of models based on the level of genetic details they included. Six levels were proposed that are still relevant:

1. Generic model with no reference to species
2. Species-specific model with no reference to cultivars
3. Genetic differences represented by cultivar specific parameters
4. Genetic differences represented by gene actions modeled through their effects on model parameters
5. Genetic differences represented by genotypes, with gene action explicitly simulated based on knowledge of regulation of gene expression and effects of gene products
6. Genetic differences represented by genotypes, with the gene action simulated at the level of interactions of regulators, gene products, and other metabolites.

Historically, ecophysiological and crop models were of Levels 1 and 2. Progressively they have included genetic information and most current models are of Level 3. Level 4 is the one currently developed. It is especially largely used to include information from quantitative genetics (outlined later in 1.2.2). Level 5 is still rarely encountered. It is restricted to the cases of model species for which the understanding of gene action in some particular physiological processes is advanced. In general, too few genes are known to feed gene-based models. However, this level may also be tested with virtual genes (example of the phenology module in GeneGro Version 2, Hoogenboom and White (2003)). Lastly, Level 6 corresponds to models simulating gene action based on interactions of regulators, gene-products, and other

metabolites. It has only been achieved in case of unicellular organisms (Tomita et al. 1999).

White and Hoogenboom (1996) and Yin et al. (2000a) were pioneers in the integration of gene action and QTL effects, respectively, in process-based models (models of Levels 4 and 5). These works were promising proofs of concept and aroused keen interest in the scientific community. The principle used in these works is to define genotypes by a set of model parameter values. These values depend both on the allelic combination carried by the genotype and on the genetic model defined from genes or QTL controlling the parameters of the model. Then the model can simulate these different genotypes.

1.3.2 QTL-based Modelling

In the absence of information on specific genes or loci, QTL analyses can be performed on model parameters. Indeed, these parameters often display quantitative and continuous variations in populations, in the same way as variables classically observed (e.g., plant height, yield, biomass). This approach was pioneered by Yin et al. (1999; 2000b), who recalculated the value of 10 genotypic parameters of the SYP-BL simulation model for barley crop growth. The major weakness of this approach was the inability of the original model to simulate observed variations. The authors suggested that the level of integration considered was not appropriate. They concluded that further physiological processes might be incorporated in the model to improve the performance of the coupling. Since, promising results have been obtained using physiological components of different traits, such as leaf elongation (Reymond et al. 2003), plant development (Yin et al. 2005; Messina et al. 2006), phenology (Nakagawa et al. 2005), early plant growth (Brunel et al. 2009), nitrogen adaptation (Laperche et al. 2006) and fruit quality (Quilot et al. 2005b). In each of these studies, QTL associated with the considered traits/processes were identified. Test of the model against independent conditions (new genotypes and new environmental conditions) also gave promising results (e.g., Reymond et al. 2003).

The latter authors focused their work on the analysis of the genetic variability of leaf elongation rate on maize in response to temperature and soil water deficit (Reymond et al. 2003, 2004). In these studies, a simple static model based on response curves of leaf elongation rate to temperature, vapour pressure deficit and soil water potential was used. Thirteen maize lines grown under six contrasted environments were used as material for validating the model, which accounted for 74 % of the genetic and environmental variations of leaf elongation rate (Reymond et al. 2004).

The QTL associated to the model parameters do not systematically co-localize with the QTL for the more integrated variables, thus highlighting the complexity of the system. For instance, no co-localization was found in maize between QTL for final leaf length under water deficit and QTL for the parameters associated to leaf expansion response to water deficit (Reymond et al. 2004).

However, co-localizations of QTL for different traits or parameters were observed in other studies. For instance, co-localizations between QTL for leaf elongation and anthesis-silking interval in maize suggest that these traits might be regulated by the same process (e.g., tissue elongation for either the leaves or the silks; Welcker et al. 2007). Similarly, the study performed by Quilot et al. (2005b) on peach gave some insight in the processes that control quality traits. Ten genotypic parameters of the virtual peach fruit model which strongly affect fruit growth and sugar accumulation (Quilot et al. 2005a) were selected among the 40 parameters of the model, for a QTL analysis (Quilot et al. 2005b). These genotypic parameters were substituted in the simulation model by the sum of QTL effects. The model was then able to account for a large part of the genetic and environmental variations in fruit size (observed and predicted values of fruit dry mass showed a correlation coefficient of 0.55). In this example, the QTL analysis of the genotypic parameters gave some insight in the processes that control quality traits, as they co-localized on the genetic map with QTL for fruit size and sugar content. This suggests putative physiological interpretations of the functions of genes under these QTL. Moreover, such results can help understand the processes involved, and thus assist the improvement of the process-based model. More details on the approach can be found in recent reviews (e.g., Hammer et al. 2006; Yin and Struik 2008; Messina et al. 2009; Bertin et al. 2010).

The number of parameters in most process-based models (from tens to hundreds) appears relatively low compared with the large number of genes of a plant (tens of thousands). Nevertheless, this simplification of the complexity of the genetic architecture shows potential anyway. Indeed, genes very often act in coordination, and it is the action of the gene group, instead of the action of each gene, that can be represented by model parameters. The set of interconnected processes controlled by such a group of genes was defined by Tardieu (2003) as "meta-mechanism". The essential is to pick out the right level of organization at the cell, organ or plant level where the consequences of the switch explaining genetic variability in the mechanism and response curves to environmental factors can still be represented to explain the observed variations of the trait of interest.

1.3.3 Gene-based Modelling

When information on specific genes is available, a gene-based model can be attempted, directly relating genotypic parameters to the expression of a few key genes. One of the earliest works was carried out by White and Hoogenboom (1996) and Hoogenboom and White (2003) using the BEANGRO simulation model for common bean (*Phaseolus vulgaris* L.). They assumed that seven genes controlled some of the genotypic parameters of the model and replaced them by linear functions describing the effect of the genes. A simple theoretical genetic model was considered with two alleles for each gene, one dominant and the other one recessive. The genotypes of 30 common bean cultivars were determined for these seven genes

and included in the model in place of the genotypic parameters. The new model (of class 5), GeneGro, simulated growth and development and could even simulate new G×E interactions. This approach has been recently included into the soybean simulation model CROPGRO-soybean to characterize the effect of six loci on growth and development, using a set of isogenic lines (Messina et al. 2006).

When a trait is controlled by a low number of major genes, modelling of gene network can also be attempted. Such an approach has been successfully used to model flowering time (Welch et al. 2003; Welch et al. 2004) and cell cycle and expansion in leaves (Beemster et al. 2006) for Arabidopsis (*Arabidopsis thaliana* (L.) Heynh.). More recently, Coen and co-workers investigated the mechanisms by which genes can control the emergence of complex shapes in *Snapdragon* flowers (Green et al. 2010) and *Arabidopsis* leaves (Kuchen et al. 2012). In their model, a set of experimentally-defined rules fixes the values of two key quantities, the direction of growth (the tissue polarity) and the local growth rates, as a function of the expression of a few genes, thus providing an implicit coupling between the cellular and the tissue levels. Petal tissue is described as a continuous sheet that grows and bends under the effect of a growth field, as specified by the tissue polarity and the local growth rates, according to elasticity theory (Kennaway et al. 2011). At the cellular level, the interactions among genes are described by a gene regulatory network that captures the evolution of gene expression levels during organ development. A specificity of the present model is that developmental changes are taken into account by explicitly modifying both the genetic control of the growth field and the regulatory interactions among genes, according to the experimental data.

Currently, the strongest limitation to develop gene-based models for complex traits is the lack of knowledge and characterization of specific genes or loci controlling these traits, including epistatic interactions and pleiotropic effects, to define the phenotypic fingerprint of cultivars for genotypic parameters. Moreover, detailed studies to quantify the environmental effects on gene expression and gene action are also required.

1.4 Modelling Cellular Networks

This section aims to draw the attention to a panel of approaches, developed in the context of systems biology that can be used to analyze, simplify and model the behaviour of cellular networks. Several formalisms are available, in a perpetual trade-off between predictive power and information needed. Here we restrict to those techniques that require a reduced amount of information as these approaches are more likely to be appropriate when dealing with large and under-characterized systems, like plants. For simplicity, we separate the analysis of gene regulatory networks and metabolic networks as they historically evolved in distinct domains, and specific tools are available.

1.4.1 Modelling and Analyzing Gene Regulatory Networks

1.4.1.1 Qualitative Models of Gene Regulatory Networks

A number of methods exist to model the dynamics of a gene regulatory network with increasing accuracy, including discrete and continuous approaches, deterministic or stochastic (see de Jong (2002) and Schlitt and Brazma (2007) for a review). Here we focus on qualitative approaches (see Appendix 1) as handy methods to capture the logic of gene control without the need for precise parameter values, rarely available for most biological systems. The simplest formalism for gene regulation is a Boolean model in which each gene is represented as binary switch that can be either on (value 1) or off (value 0). At any given time, the state of the network is represented by the n-tuple of zeros and ones describing which gene is active or inactive. Transitions from one state to another are determined by gene mutual regulation, expressed by means of logical rules. The dynamics of a logical network is thus represented by a sequence of states, describing all possible activation/inactivation pathways compatible with a given regulatory logic. Any trajectory naturally leads to a steady state, i.e., an expression pattern that is maintained indefinitely by the system (it can be either a fixed point or a cycle). The existence of multiple steady states is often associated with the existence of distinct developmental outcomes.

The disadvantage of Boolean models resides in their high level of abstraction that makes it difficult to integrate (when available!) data on promoter activity and sequential gene regulation. Indeed, a same transcription factor can regulate the expression of several genes depending on its concentration. In this perspective, an extension to the Boolean formalism is the logical models, in which the simple binary nature of gene activation/inactivation is replaced by a variable that is able to assume p discrete values (0, 1, 2, …, p). In this way, a level-dependent action of each transcription factor can be included into the logical rules (see Fig. 1.1 for an example). Alternatively to logical models, piece-wise linear models offer a formalism that is somehow closer to a continuous description, combining the qualitative approach with a time-continuous description of gene regulation (see Appendix 1 for more information). We refer to Snoussi (1989) and Wittmann et al. (2009) for a discussion on the relation between logical and piece-wise linear models. For both types of models, software is available for the construction and analysis of a qualitative model of gene regulatory networks (de Jong et al. 2003; Gonzalez et al. 2006).

Thanks to their simplicity, qualitative approaches, in combination with formal verification tools (Monteiro et al. 2007) can be used to test the coherence of experimental data, possibly pointing out missing regulators or interactions between components.

In this perspective, a promising approach is the one proposed by Li et al. (2006) to investigate the functional basis of Abscisic Acid (ABA) signaling. Starting from sparse literature data, including protein-protein interactions, knockout experiments and pharmacological tests, the authors developed an heuristic qualitative reasoning to (a) assemble a consistent signal transduction network of ABA-induced stomatal closure and (b) build a (Boolean) dynamical model of the system. Interestingly, the

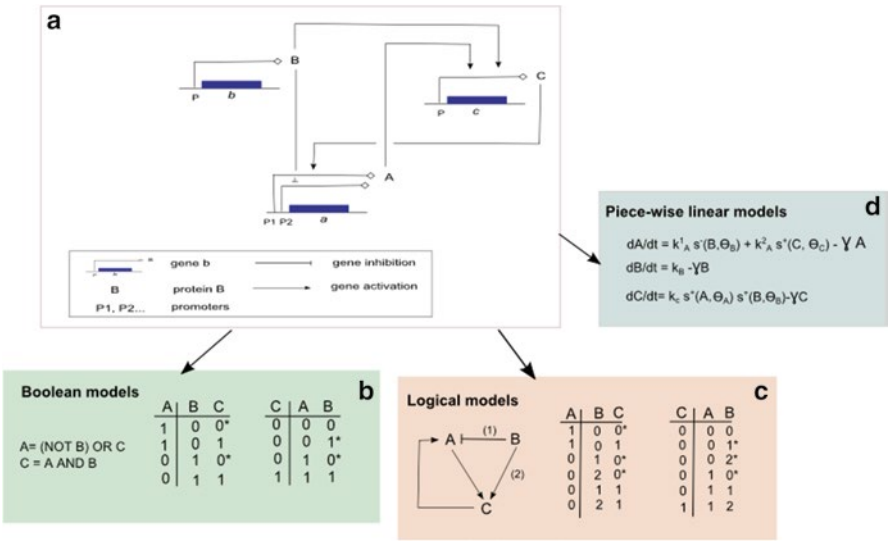


Fig. 1.1 Example of qualitative modelling of a gene regulatory network using a Boolean, logical and piece-wise linear formalism. (**a**): example of gene regulatory network. (**b**): Boolean equations and corresponding table of truth, for network in panel (**a**). Boolean operators OR, AND and NOT are used to describe the logic of gene interactions. The steady states of the system are indicated with a star. (**c**): Graph representation of the network in *A*, where the edges are labelled to express the rank number of the interaction, and corresponding table of truth in the logical formalism. (**d**): Piece-wise linear equations corresponding to the network in panel *A*. k_A, k_B and k_C are synthesis rates, γ_A, γ_B and γ_C are degradation rates and θ_A, θ_B, and θ_C are threshold concentrations for proteins A, B and C, respectively

lack of quantitative information on process kinetics is circumvented by randomly sampling all possible updating orders (Chaves et al. 2006) and computing an average probability of stomatal closure over 10,000 initial conditions. The model is then used to predict essential components of the system, evaluating the effects of single and multiple node disruptions on the resulting responsiveness of stomatal closure. Notice that due to its qualitative nature, model predictions are valid independently of the specific kinetic properties of the system. Nevertheless, future quantitative data on biochemical mechanisms could be easily accounted for, by replacing the corresponding Boolean rule with a stochastic or continuous description, in the framework of hybrid modelling approach (Chaouiya 2007; Chaves et al. 2006).

1.4.2 Modelling and Analyzing Metabolic Networks

1.4.2.1 Steady-state Models of Metabolic Networks

The so-called Constraint-Based (CB) approaches use the stoichiometry and the thermodynamics of biochemical reactions as mathematical constraints to progressively reduce the space of possible steady-state solutions of the metabolic system of

equations (see Appendix 2). In this framework, Metabolic Pathways Analysis aims at describing all possible steady-state behaviours of the system, compatible with mass balance and thermodynamics as constraints (Papin et al. 2003; Schilling et al. 2000; Schuster et al. 2000). The aim here is to provide a general characterization of network capabilities, pointing at specific designing features that may provide insights into the functional organization of the system (Rios-Estepa and Lange 2007; Stelling et al. 2002).

Among all possible behaviours predicted by pathway analysis, only few are actually realized by a given organism, depending on environmental conditions. The hypothesis behind this observation is that a selective-external pressure acts as an additional constraint that favors few specific flux distributions instead of others. Flux Balance Analysis (FBA) is a method developed by Palsson and collaborators that aims to predict such reasonable (i.e., likely to be realized) flux distributions by assuming that they maximize a given objective function (Orth et al. 2010). Recently, a number of improvements to FBA have been developed, with the aim of relaxing the notion of optimality. The impact of alternative optimal states (i.e., alternative solutions that share the same optimal score) have been analyzed (Mahadevan and Schilling 2003) as well as the existence of non-optimal solutions (Mahadevan and Schilling 2003; Segrè et al. 2002), (i.e., solutions that score near the optimal value but not exactly). These solutions indeed may be more appropriate to describe the behaviour of genetically engineered organisms (e.g., a knockout mutation) or to compare with experimental data, for which optimal growth conditions are not guaranteed.

In plants, a number applications of pathways analysis and FBA are starting to appear, as reviewed by Rios-Estepa and Lange (2007) and Sweetlove and Ratcliffe (2011). In spite of these successes, two main problems currently limit the application of CB methods to plants. The first is the presence of sub-cellular compartments that can modify the predicted flux distribution when not a priori included in the metabolic model. To date, assigning a reaction to a specific compartment is far from trivial. Experimental data are still scarce and information on metabolite transport mechanisms between different organelles is often lacking, due to technical difficulties (Allen et al. 2009). Further advancements in Nuclear Magnetic Resonance spectroscopy and fluorescence methods however are meant to overcome these problems, contributing to the development of more realistic CB models. The second issue, proper to FBA, is more fundamental and regards the maximization of an appropriate objective function in plants (Sweetlove and Ratcliffe 2011). Historically, FBA have been developed for microbes for which good objectives functions may be established (generally optimal biomass production) and whose rapid adaptive evolution guarantees a near-optimal functioning, soon after a perturbation. In the case of plants, adaptation is generally too slow to guarantee optimality in a reasonable experimental time and the choice of maximal cost function can be called into question. In this context, alternative methods using sub-optimal solution may be more appropriate. The method of minimization of metabolic adjustment (Segrè et al. 2002) in particular has been suggested as especially suitable as it allows to investigate the effect of a perturbation starting by the knowledge of the initial flux distribution only. Experimentally determined flux distribution (via steady-state

isotope-labeling) may therefore be used, thus avoiding the problem of the objective function definition.

1.4.2.2 Dynamical Models of Metabolic Networks

Plants, more than other systems, are continuously subject to fluctuations of environmental factors that can induce rapid rearrangement in metabolite levels or in metabolic fluxes. In this situation, the steady-state assumption behind constraint-based methods can become a limit. Further insights into the dynamics of metabolism and in the "what-if" scenarios can be obtained by building a kinetic model of the system, describing how metabolites concentrations change over the time due to the interactions between other molecules. Traditionally, the dynamics of a metabolic system is described as a set of nonlinear ordinary differential equations (ODE system) assigning to each reaction a rate law (e.g., Michaelis-Menten kinetics, as a classical choice) which describes how its speed depends on the concentration of other molecules (metabolites, enzymes). The choice of the appropriate rate law and its complete definition is extremely expensive and requires a good knowledge of all biochemical steps. For this reason, most available kinetic models describe only small metabolic networks (a dozen of variables) where kinetic information and parameters values have been derived from literature data (Uys et al. 2007; Nägele et al. 2010) or from dedicated experiments (Curien et al. 2009; Beauvoit et al. 2014).

For larger networks, such information is generally not available and approximated expressions for reaction rates have to be used.

A first strategy is based on the use of simplified kinetics that are valid near a reference state, usually chosen to be the steady state of the system (Heijnen 2005; Stitt et al. 2010). Smallbone et al. (2007) propose a method based on Flux Balance Analysis. The idea is to make fluxes vary according to a lin-log kinetics (Visser and Heijnen 2003; Heijnen 2005) around their steady state value (as determined by FBA) and then predict the evolution of metabolic concentrations according to the ODE system. The main advantage of this approach resides in the limited information required; indeed, the model can be defined using the information contained in the stoichiometric matrix only, even in the absence of experimental data for kinetics parameters. Of course, limitations already discussed for FBA naturally apply to this approach too, especially in what concerns the assumption of optimality for complex systems like plants.

A step further along the simplification of reaction rates, qualitative methods offer a way to investigate essential dynamical properties of the network that are invariant for a range of reaction mechanisms and parameters values. Among all available formalisms for qualitative modelling, Petri Nets (see Appendix 1) are particularly suitable for metabolism because their structure agrees well with the idea of conversion embodied in biochemical reactions (Chaouiya 2007). In particular, Petri Nets allow taking into account reaction stoichiometry and differences in reaction rates ("delay"), when this information is available.

In the case of well-studied systems, Palsson and coworkers proposed an alternative approach for modelling large metabolic networks that relies on the increasing availability of high-throughput data (Jamshidi and Palsson 2008b, 2010). To this aim, a stoichiometric model of the network is combined to a mass action description of rate laws. Data on flux distribution, metabolites concentration and equilibrium constants are used to estimate kinetic parameters and simulate the dynamics of the system around a particular steady state, for which data are available. Regulatory effects (enzymes binding, allosteric regulation) can also be taken into account by directly modelling the regulator-substrate binding reaction, once information on enzymes concentrations and binding rates are available. The advantages of this method reside in its scalability, as demonstrated by its application to the red-blood cell metabolism (Jamshidi and Palsson 2008b; Kauffman et al. 2002), and in the possibility of automatically refining the model, as long as new omics data are collected. The need for a rich and reliable set of experimental data, however, currently limits its application to plants.

1.5 Integrating Cellular Networks into Plant Models

When biological information is available, a mechanistic description of cellular networks can be attempted and integrated into plant models, in the perspective of 'crop systems biology' (Yin and Struik 2010). Such integration is relevant in two ways. From an ecophysiological point of view, the integration of cellular and molecular levels can help to refine plant models, shedding light onto the complex interplay between different spatial and temporal scales in the emerging system response. In particular, the presence of explicit molecular variables can help to identify those molecular mechanisms that may convey interesting agronomical properties to current crops varieties. From the point of view of molecular biology, the existence of an integrated model could offer a useful framework for interpreting omics data, in relation to environmental factors and agricultural practices.

The ambition of the so-called multi-scale models is to explicitly integrate mechanisms that take place on different temporal or spatial scales, while keeping the computational cost low (Baldazzi et al. 2012; Southern et al. 2008). Two main issues characterize these models: (1) the (simplified) description of processes on a common scale, and (2) the way different scales are connected, i.e., how the information is passed among organizational levels.

1.5.1 Model Simplification

In order to be coupled, models on a single scale must be reasonably simple to avoid computational problems but, at the same time, sufficiently complex to correctly represent the expected biological behaviour. The claim is that, for most

ecophysiological questions, there is no need for detailed modelling: at the scale of plant development and adaptation, only few molecular mechanisms and variables are likely to significantly affect the behaviour of the system, and need to be explicitly accounted for (Hammer et al. 2006; Génard et al. 2007). Here we review a few of methods that can be used to analyze cellular networks and obtain a simplified representation of cellular functioning, in both variable number and mathematical expression.

1.5.1.1 Structural Analysis

As a preliminary step, the inspection of network topology by means of statistical and graph analysis methods can provide useful insights into the regulatory architecture of the system (see Barabási and Oltvai (2004) for a review), pointing at few nodes that "naturally" emerge as key variables of the system. The analysis of transcriptional regulatory networks in unicellular systems (Barabási and Oltvai 2004; Ma et al. 2004a) but also eukaryotic systems (Carrera et al. 2009) for instance, has uncovered a typical hierarchical structure, with few genes (hubs) having a huge number of outgoing connections (i.e., regulating a large number of genes). These genes thus represent a sort of 'master' regulators of the network, able to control most biological functions (Martínez-Antonio and Collado-Vides 2003; Seshasayee et al. 2009) and their adaptation to environmental changes (Görke and Stülke 2008; Hengge-Aronis 1999).

Another important aspect of structural analysis is the search for functional modules, i.e., sub-networks able to work (almost) independently of the rest of the network (Wagner et al. 2007). Several methods try to identify modules in an automated way (Wang et al. 2008). Most of them make use of connectivity properties (Ma et al. 2004b; Schuster et al. 2002; Tanay et al. 2004) whereas others combine topology and experimental data to increase their interpretability in terms of biological function (Mao et al. 2009; Sridharan et al. 2012). In plants modular organization has been recently highlighted in Arabidopsis (Mao et al. 2009): a hundred of modules have been identified and assigned to the main biological processes. Some of these modules are large (>1000 genes), like the one related to photosynthesis but smaller modules are also identified for specific processes, as the one related to starch metabolism that includes only 10 genes, thus providing a reasonable starting point for further characterization.

1.5.1.2 Time Scale Analysis

When integrating processes of different nature, an analysis of typical time-scales involved can prove extremely useful for model reduction.

Based on time-scale separation, the original model can be usually rewritten into (at least) two distinct subsystems, corresponding to slow processes and fast processes:

$$dx_s/dt = f_s\left(x_s, x_f, p\right)$$
$$dx_f/dt = f_f\left(x_s, x_f, p\right)$$

where x_s *and* x_f are the slow and fast variables, respectively, and p are parameters.

If the time-scales of the processes differ by at least an order of magnitude a few approximations are likely to be possible. Variables that are changing on a time-scale much slower than the one of interest can simply be assumed constant, averaging out (small) variations in the time-window of interest (Radulescu et al. 2008). For variables that are evolving on a time-scale much faster than the reference time-scale, one commonly assumes the fast variables to be in a quasi-steady-state, i.e., instantly adapting to changes occurring on the reference time-scale. Mathematically speaking, this means that the dimension of the model can be reduced by setting the time derivative of the fast system to zero (Heinrich and Schuster 1996), thus resulting in a simple set of algebraic equations.

$$dx_s/dt = f_s\left(x_s, x_f, p\right)$$
$$dx_f/dt = f_f\left(x_s, x_f, p\right) = 0$$

If a kinetic model of the system is available, the analysis of time-scales can be rigorously performed by means of modal analysis (Jamshidi and Palsson 2008a). The application of this method naturally leads to pooling of variables into groups of species that evolve in a coordinated fashion above a specific time-scale. This means that the model size can be effectively reduced by considering the dynamics of pools as representative of the dynamics of their constitutive species. Differences in the typical time-scales of metabolism (of the order of seconds) and gene expression (minutes to hours) have been recently exploited to investigate the coupling between metabolic and genetic networks in bacteria (Baldazzi et al. 2010; Covert et al. 2008; Shlomi et al. 2007).

1.5.1.3 Metabolic Control Analysis

When dealing with metabolic networks, the stoichiometry of the biochemical reactions imposes a rigid constraint to the dynamics of the system: any change in a flux (or metabolite concentration) must be compensated by other changes in the network, thus linking local kinetics properties to the global system behaviour.

Metabolic control analysis (MCA) is a strategy to analyse how the control of a metabolic pathway is shared among the different reactions (Heinrich and Rapoport 1974; Kacser and Burns 1973). The degree of control of a reaction is quantified in terms of control coefficients, defined as the fractional change of the system property

(flux (ν) or metabolite concentration (x), at steady state) in response to a change in enzyme activity E.

$$\text{Flux control coefficient: } C_{ij}{}^{v} = \frac{\partial v_i}{\partial E_j} \cdot \frac{E_j}{v_i} = \frac{\partial \ln v_i}{\partial \ln E_j}$$

$$\text{Concentration control coefficient: } C_{ij}{}^{x} = \frac{\partial x_i}{\partial E_j} \cdot \frac{E_j}{x_i} = \frac{\partial \ln x_i}{\partial \ln E_j}$$

A zero control coefficient means that the system variable does not change when enzyme activity is modified. A flux control coefficient of 1 means that the reaction catalysed by the enzyme completely determines the flux value. In these conditions, no other reaction can affect the flux and the enzyme is said to be rate-limiting. In nature, rate limiting reactions are very rare. System control is generally shared among multiple steps and MCA allows ranking their importance, once the target variable has been defined.

Within the context of model reduction, MCA can be used to identify those enzymes that most affect a target process and that should therefore be retained in the model. Geigenberger et al. (2004), for instance, used MCA to investigate the biosynthetic pathway of starch in potato tubers, showing that starch accumulation is mostly controlled at the level of ATP transport between cytosol and amyloplast, with a minor role for starch synthesis enzymes.

1.5.2 Coupling Among Scales

Once the description of processes on a common scale has been defined, models have to be connected together. The scaling up from cell to tissue or organs properties implies understanding the way cells communicate and coordinate together. Recent studies have shown that cell-to-cell coupling may involve different but intertwined mechanisms that include biochemical signalling as well as physical processes, as in the case of relaxation of mechanical stresses (Howard et al. 2011). A full multi-scale approach requires the identification of the predominant mechanisms and of those molecular variables that effectively act as "hubs", connecting different organizational levels (Keurentjes et al. 2011).

This is the case of calcium ions in heart models, probably the most advanced example of multi-scale approach (Hunter and Borg 2003; Noble 2002). The calcium intracellular concentration indeed affects the kinetics of actin filaments, providing the desired link between cellular metabolism and local mechanical properties of the heart. At the cellular scale, a set of ODE equations describe the temporal change of local concentration of ions (calcium and sodium mainly) due to the presence of ion channels, ion pumps and exchangers etc. The stretching of actin fibres, following Ca^{2+} binding, is modelled using a first-order kinetics and the fibre tension computed as a function of the intracellular calcium concentration. A diffusion equation, solved

by means of a finite elements method, then describes the propagation of the mechanical wave on the scale of the whole organ. Interestingly, the organ level can also exert a feedback to the cellular one. In the model by Hunter et al. (1998) mechanical perturbation of the heart can alter the release of calcium ions from specific regulatory proteins, thus affecting the intracellular Ca^{2+} concentration.

In plants, multi-scale approaches are often employed to investigate organ emergence and morphogenesis, in both vegetative and non-vegetative organs (Band et al. 2012a; Prusinkiewicz and Runions 2012). A multi-scale model has been recently proposed to explain the dynamics of cell elongation, in *Arabidopsis thaliana* roots and its control by the hormone gibberellin (Band et al. 2012b). To this aim, root elongation zone is described as a single cell file along which gibberellin can diffuse. The movement of gibberellin hormone is described in details, both within and across the cells, as well as cell expansion along the elongation zone. At the subcellular level, a complex signalling network links the concentration of gibberellin to the distribution of DELLA proteins, a known growth repressor. Following vacuole expansion, gibberellin is rapidly diluted in the cell creating a significant concentration gradient of both hormone and DELLA proteins along the elongation zone. The model predicts indeed a progressive accumulation of DELLA towards the end of the elongation zone, thus explaining the observed reduction of cell elongation.

The above examples illustrate the potential of multi-scale modelling in elucidating the emergence of complex phenotypes. Such examples however are still rare. Our knowledge of biological systems and plants in particular remains limited and further efforts are needed to understand the interplay among different organizational levels and identify potential hubs. At term, interactions with environment should also be taken into account in order to develop a full multi-scale ecophysiological model.

1.6 Conclusions, Open Issues and Perspectives

Recent advancement in high-throughput methods has led many authors to hope for the advent of crop systems biology, combining information from molecular biology and physiology with a broader look to plant development and growth, in relation to environmental factors and agricultural practices. This is a great challenge of integrative biology that needs the collaborative work of many disciplines. The success of crop systems biology relies on the close collaboration of scientists from different fields (including biology and mathematics), with several iterative cycles between experiments and models.

From a modelling point of view this calls for the development of new methods to account for gene control. Classical approaches by parameter specification have already proven their various interests. Among them we can highlight that integration to crop models is direct and does not require rebuilding new specific models and that quantitative genetic control can easily be taken into account. The drawback of course is a low explicative power of the fine mechanisms; model predictions may be

valid only in well-defined experimental conditions, making this kind of model less amenable to exploratory *in silico* research. In parallel, approaches have been developed to model the behaviour of cellular networks. At this organization scale, knowledge is more complete and specific genes or loci controlling the traits are better known and described. A panel of approaches are available, but in this chapter we focused on gene regulatory networks and metabolic networks that could then be linked to crop models. The last part of the chapter was devoted to the integration of different temporal and spatial scales (cellular, tissue, organ) within a single model, in the perspective of multi-scale modelling.

Plant and crop ecophysiological models formalize traits as the result of genotypic and environmental effects and the relations among traits. In this sense they provide a platform of virtual profiling (i) for the integrative analyses of the impact of a combination of traits on whole plant and crop phenotype (e.g., Bertin et al. 2010; Hammer et al. 2009) and (ii) for quantifying individual impact of traits, or in interaction with other traits in a trait network, within a range of agro-climatic conditions. This opens the door towards new opportunities of virtual breeding of ideotypes such as developing genotypes specifically adapted to a set of conditions of particular interest (e.g., non-optimal pedoclimatic scenarios, new cultivation techniques, future climates). However, going back to experiments will be crucial to assess simulations by experimental evidences (Andrivon et al. 2013).

Despite recent advances in knowledge and development of tools, several scientific and technical challenges still need to be overcome. Considerable efforts are still needed to make the links between levels of organization and to integrate different types of information but we believe that this framework will soon become essential to further decipher plant complexity.

Appendix 1: Qualitative Modelling

Qualitative approaches aim to characterize key properties of the dynamics of the system as the existence of steady states, limit cycles (oscillation) but also specific dynamical patterns, like for instance the subsequent activation of two network components. Due to their qualitative nature, these approaches are generally non-deterministic: the result of a qualitative simulation is a state transition graph containing all possible dynamical pathways that are coherent with the initial model definition. In this perspective, model checking tools provide an useful complement, allowing the automated verification of dynamical properties of the system, regardless of the specific trajectory (Monteiro et al. 2007).

In the following we present the basics of three common qualitative models; the interested reader can find relevant references for more information.

Logical and Boolean Model

Variables (usually genes or signaling molecules) are represented as discrete entities that can assume a set of integer values (0 and 1 in the Boolean case, 0,1,2, … in logical models) describing which variable is present and, in the case of logical models, at which concentration (state of the system). The temporal evolution of the system, i.e. the transition between system states, is determined by the interactions among variables. A steady state is reached when the output state is equal to the input one. A combination of logical functions (AND, OR, NOT) or, alternatively, a table of truth can be used to define the interaction rules. In the case of logical models, the interaction between two variables can be submitted to an additional constraint on the concentration of the regulatory variables. So for instance, in the example in Fig. 1.1 gene C is activated by protein B if and only if the latter is present at high concentration (variable level 2). The dynamics of Boolean and logical models (but not their steady states) depends on the specific updating scheme employed. Two main kinds of update are possible, either synchronous or asynchronous. In the synchronous scheme, all nodes are updated simultaneously according to their values at previous time, whereas in the asynchronous update a single node, selected at random, is changed at time, according to the current state of the network. For more information on logical models and their application, see de Jong (2002), Fauré et al. (2006) and Morris et al. (2010).

Piece-wise Linear (PL) Model

PL model, introduced by Glass and Kauffman (1973), lies in between ODE and logical models (Davidich and Bornholdt 2008): a continuous step function describes the activation (inhibition) of a gene, whenever the concentration of a regulatory variable x_j is above given threshold Θ_j (Fig. 1.2).

$$s^{+}\left(xj,\theta j\right)=\begin{cases}1 & xj>\theta j\\ 0 & xj<\theta j\end{cases}$$

$$s^{-}\left(xj,\theta j\right)=1-s^{+}\left(xj,\theta j\right)$$

Threshold concentration thus defines a rectangular partition of the phase space, such that in every region not located on a threshold, the PL model reduces to a linear system of differential equations. Moreover, in every such region the derivatives

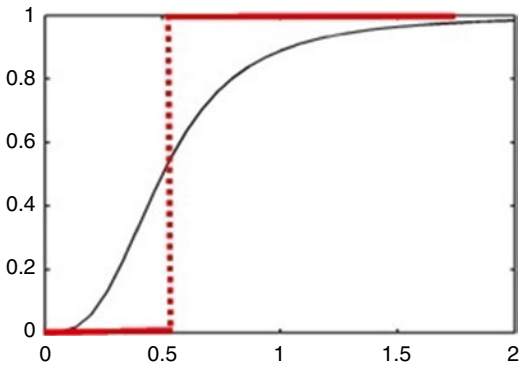


Fig. 1.2 Piece-wise linear approximation of a sigmoid function into a step function

(trends) of the concentration variables have a determinate sign, which is shown to be invariant for rather weak constraints (inequalities) on the parameter values. In a qualitative analysis, each of these regions corresponds to a qualitative state of the network, analogous to the n-tuple of zero and ones of Boolean models.

PL models are often used to describe gene regulatory networks, based on the observation that gene expression rates are often a sigmoid function of the transcription factor concentration. Step-functions are thus seen as an approximation of this sigmoidal behaviour. For more information on PL formalism and its application to biological systems see Baldazzi et al. (2011) and de Jong et al. (2004).

Petri Net Model

Petri nets contain two kinds of nodes: places and transitions. Places (graphically described as circle) represent the resources or variables of the system (e.g., metabolites in a metabolic network) whereas transitions (rectangles) correspond to events (e.g., reactions) that can change the state of the variables. At any time, each place contains a zero or positive number of tokens (small black dots): when the number of tokens is sufficient, the transition is enabled and the reaction can take place. The firing of an enabled transition results in the consumption of tokens in the input place and the creation of a given number of tokens in the output place, according to edge weight (Fig. 1.3). Due to their structure, Petri nets are usually employed to describe signalling and metabolic networks. Indeed, following a metabolic analogy, the number of tokens in a given place can be associated with metabolite concentrations whereas edge weights correspond to reaction stoichiometry. For a more complete review we refer to Chaouiya (2007).

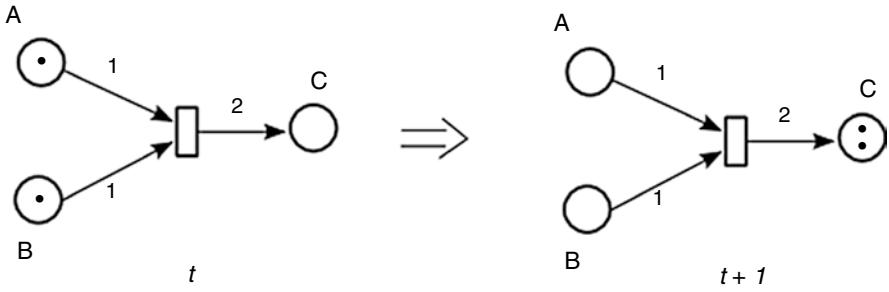


Fig. 1.3 Firing of a Petri net. At time t, places A and B both contain one token (*small black dot*), enough to allow the reaction A + B–>2C to take place, as indicated by the edge weights (*small numbers on the* arrows). At time t + 1, tokens in A and B have been consumed whereas two tokens have been created in place C

Appendix 2: Constraint-based Models

Metabolic Pathway Analysis

The idea is to study the solution space (flux vector v) of the following system at steady-state:

$$N \cdot v = 0$$
$$v \in C$$

where N is the stoichiometric matrix and *C* is the set of constraints. In the simplest version *C* are thermodynamics constraints, defining irreversible reactions (i.e., vi > 0 for some flux i), but it can also include flux capacity constraints (i.e., v_min < vi < v_max) or experimental data (measured flux values, i.e., vi = mi). In this framework, omics data can be used directly to reduce the search area (in blue in Fig. 1.4).

Within constraints *C*, any admissible flux distribution is represented as a non-negative combination of a valid set of pathways that is (a) unique and (b) minimal, for each network topology. In particular each pathway is "non decomposable", i.e., it consists of the minimum number of reactions needed to exist as a functional unit. Two alternative definitions of these pathways exist, elementary modes (Stelling et al. 2002) and extreme pathways (Schuster et al. 2000), that can provide slightly different information (see Papin et al. (2004) for a comparison between the two).

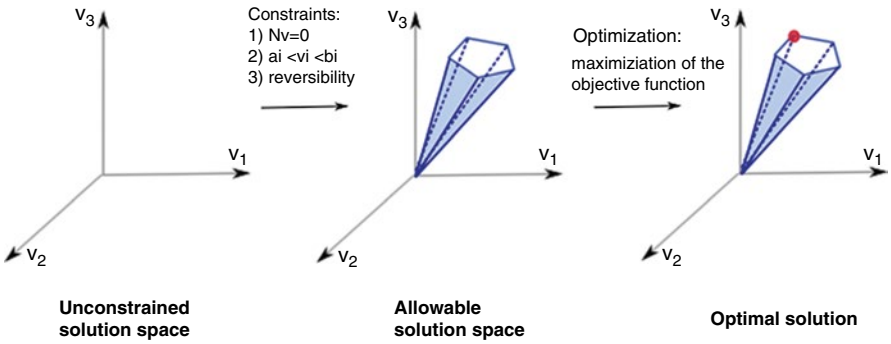


Fig. 1.4 Schematic representation of Constraint-Based Analysis of metabolic network (inspired by Orth et al. (2010)). In absence of constraints, the flux distribution of a metabolic network can be everywhere in the solution space. Stoichiometry, thermodynamics and limitations to fluxes capacity act as constraints that reduce the allowable space of solutions (*blue surface*). An optimization principle allows selecting one optimal solution, at one edge of the solution space

Flux Balance Analysis and Related Techniques

Within the solution space defined by the Metabolic Pathway Analysis, Flux Balance Analysis aims at identifying the most likely flux distribution as the ones that maximize a given objective Z, i.e.:

$$N \cdot v = 0$$
$$v \in C \mid Z(v)\max$$

Several choices are possible regarding the objective function Z and the formulation of an appropriate one is still subject of research, especially in the case of higher organisms (Schuetz et al. 2007). Among the most popular cost functions in literature are the biomass or ATP productions, growth rate or the production of specific metabolites.

Whatever the objective function, a common problem with FBA is the possible existence of multiple optimal solutions, i.e. several flux distributions with same cost score. In this case the prediction is not unique, and some conclusions may change depending on the selected solution. However, alternative solutions may have a biological meaning: biological systems show a high degree of redundancy that is often associated with a certain functional robustness. Mahadevan and Schilling (2003) explicitly use the existence of multiple optima to investigate redundancies in the network and to derive a flux range in which the optimality is guaranteed. In the case of engineered organisms (e.g., mutants or knock-outs), MOMA technique predicts the expected flux distribution by assuming that it is the closest to the wild type one (principle of minimal modification), irrespectively of its optimality (Segrè et al. 2002).

References

Allen D, Libourel I, Shachar-Hill Y (2009) Metabolic flux analysis in plants: coping with complexity. Plant Cell Environ 32:1241–1257. doi:10.1111/j.1365-3040.2009.01992.x

Andrade-Sanchez P, Gore MA, Heun JT, Thorp KR, Carmo-Silva AE, French AN, Salvucci ME, White JW (2013) Development and evaluation of a field-based high-throughput phenotyping platform. Funct Plant Biol 41:68–79. doi:10.1071/FP13126

Andrivon D, Giorgetti C, Baranger A, Calonnec A, Cartolaro P, Faivre R, Guyader S, Lauri PE, Lescourret F, Parisi L, Ney B, Tivoli B, Sache I (2013) Defining and designing plant architectural ideotypes to control epidemics? Eur J Plant Pathol 135:611–617. doi:10.1007/s10658-012-0126-y

Baldazzi V, Ropers D, Markowicz Y, Kahn D, Geiselmann J, de Jong H (2010) The carbon assimilation network in Escherichia coli is densely connected and largely sign-determined by directions of metabolic fluxes. PLoS Comp Biol 6, e1000812

Baldazzi V, Monteiro PT, Page M, Ropers D, Geiselmann J, de Jong H (2011) Qualitative analysis of genetic regulatory networks in bacteria. Understanding the dynamics of biological systems. Springer, Dordrecht

Baldazzi V, Bertin N, de Jong H, Génard M (2012) Towards multiscale plant models: integrating cellular networks. Trends Plant Sci 17:728–736. doi:10.1016/j.tplants.2012.06.012

Band LR, Ja F, Godin C, Jensen OE, Pridmore T, Bennett MJ, King JR (2012a) Multiscale systems analysis of root growth and development: modeling beyond the network and cellular scales. Plant Cell 24:3892–3906. doi:10.1105/tpc.112.101550

Band LR, Ubeda-Tomás S, Dyson RJ, Middleton AM, Hodgman TC, Owen MR, Jensen OE, Bennett MJ, King JR (2012b) Growth-induced hormone dilution can explain the dynamics of plant root cell elongation. Proc Natl Acad Sci U S A 109:7577–7582. doi:10.1073/pnas.1113632109

Barabási A-L, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. Nat Rev Genet 5:101–113. doi:10.1038/nrg1272

Beauvoit BP, Colombié S, Monier A, Andrieu MH, Biais B, Bénard C, Chéniclet C, Dieuaide-Noubhani M, Nazaret C, Mazat JP, Gibon Y (2014) Model-assisted analysis of sugar metabolism throughout tomato fruit development reveals enzyme and carrier properties in relation to vacuole expansion. Plant Cell 26(8):3224–3242. doi:10.1105/tpc.114.127761

Beemster GTS, Vercruysse S, De Veylder L, Kuiper M, Inzé D (2006) The arabidopsis leaf as a model system for investigating the role of cell cycle regulation in organ growth. J Plant Res 119:43–50

Bermudez L, Urias U, Milstein D, Kamenetzky L, Asis R, Fernie AR, Van Sluys MA, Carrari F, Rossi M (2008) A candidate gene survey of quantitative trait loci affecting chemical composition in tomato fruit. J Exp Bot 59:2875–2890

Bertin N, Martre P, Genard M, Quilot B, Salon C (2010) Under what circumstances can process-based simulation models link genotype to phenotype for complex traits? Case-study of fruit and grain quality traits. J Exp Bot 61:955–967. doi:10.1093/jxb/erp377

Blanco A, Pasqualone A, Troccoli A, Di Fonzo N, Simeone R (2002) Detection of grain protein QTLs across environments in tetraploid wheats. Plant Mol Biol 48:615–623

Börner A, Worland AJ, Plaschke J, Schumann E, Law CN (1993) Pleiotropic effects of genes for reduced height (rht) and day-length insensitivity (ppd) on yield and its components for wheat grown in middle Europe. Plant Breed 111:204–216

Brunel S, Teulat-Merah B, Wagner MH, Huguet T, Prosperi J, Dürr C (2009) Using a model-based framework for analysing genetic diversity during germination and heterotrophic growth of Medicago truncatula. Ann Bot 103:1103–1117. doi:10.1093/aob/mcp040

Carrera J, Rodrigo G, Jaramillo A, Elena S (2009) Reverse-engineering the Arabidopsis thaliana transcriptional network under changing environmental conditions. Genome Biol 10:R96. doi:10.1186/gb-2009-10-9-r96

Causse M, Duffe P, Gomez MC, Buret M, Damidaux R, Zamir D, Gur A, Chevalier C, Lemaire-Chamley M, Rothan C (2004) A genetic map of candidate genes and QTLs involved in tomato fruit size and composition. J Exp Bot 55:1671–1685

Causse M, Chaïb J, Lecomte L, Buret M, Hospital F (2007) Both additivity and epistasis control the genetic variations for fruit quality traits in tomato. Theor Appl Genet 115:429–442

Chaïb J, Lecomte L, Buret M, Causse M (2006) Stability over genetic backgrounds, generations and years of quantitative trait locus (QTLs) for organoleptic quality in tomato. Theor Appl Genet 112:934–944

Chaouiya C (2007) Petri net modelling of biological networks. Brief Bioinform 8:210–219

Chaves M, Sontag ED, Albert R (2006) Methods of robustness analysis for Boolean models of gene control networks. IEE P Syst Biol 153:154–167

Covert M, Xiao N, Chen T, Karr J (2008) Integrating metabolic, transcriptional regulatory and signal transduction models in Escherichia coli. Bioinformatics 24:2044–2050. doi:10.1093/bioinformatics/btn352

Curien G, Bastien O, Robert-Genthon M, Cornish-Bowden A, Cárdenas ML, Dumas R (2009) Understanding the regulation of aspartate metabolism using a model based on measured kinetic parameters. Mol Syst Biol 5:271

Davidich MI, Bornholdt S (2008) Boolean network model predicts cell cycle sequence of fission yeast. PLoS One 3:e1672. doi:10.1371/journal.pone.0001672

de Jong H (2002) Modeling and simulation of genetic regulatory systems: a literature review. J Comp Biol 9:67–103. doi:10.1089/10665270252833208

de Jong H, Geiselmann J, Hernandez C, Page M (2003) Genetic network analyzer: qualitative simulation of genetic regulatory networks. Bioinformatics 19:336–344

De Jong H, Geiselmann J, Batt G, Hernandez C, Page M (2004) Qualitative simulation of the initiation of sporulation in Bacillus subtilis. B Math Biol 66:261–299. doi:10.1016/j.bulm.2003.08.009

Dudley JW, Clark D, Rocheford TR, LeDeaux JR (2007) Genetic analysis of corn kernel chemical composition in the random mated 7 generation of the cross of generations 70 of ihp×ilp. Crop Sci 47:45–57

Fauré A, Naldi A, Chaouiya C, Thieffry D (2006) Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. Bioinformatics 22:e124–e131. doi:10.1093/bioinformatics/btl210

Geigenberger P, Stitt M, Fernie AR (2004) Metabolic control analysis and regulation of the conversion of sucrose to starch in growing potato tubers. Plant Cell Environ 27:655–673. doi:10.1111/j.1365-3040.2004.01183.x

Génard M, Bertin N, Borel C, Bussières P, Gautier H, Habib R, Léchaudel M, Lecomte A, Lescourret F, Lobit P et al (2007) Towards a virtual fruit focusing on quality: modelling features and potential uses. J Exp Bot 58:917–928

Glass L, Kauffman SA (1973) The logical analysis of continuous, non-linear biochemical control networks. J Theor Biol 39:103–129

Gonzalez AG, Naldi A, Sánchez L, Thieffry D, Chaouiya C (2006) GINsim: a software suite for the qualitative modelling, simulation and analysis of regulatory networks. Biosystems 84:91–100. doi:10.1016/j.biosystems.2005.10.003

Görke B, Stülke J (2008) Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. Nat Rev Microbiol 6:613–624. doi:10.1038/nrmicro1932

Granier C, Aguirrezabal L, Chenu K, Cookson SJ, Dauzat M, Hamard P, Thioux J-J, Rolland G, Bouchier-Combaud S, Lebaudy A, Muller B, Simonneau T, Tardieu F (2006) PHENOPSIS, an automated platform for reproducible phenotyping of plant responses to soil water deficit in Arabidopsis thaliana permitted the identification of an accession with low sensitivity to soil water deficit. New Phytol 169:623–635. doi:10.1111/j.1469-8137.2005.01609.x

Green A, Kennaway JR, Hanna A, Bangham JA, Coen E (2010) Genetic control of organ shape and tissue polarity. PLoS Biol 8:e1000537. doi:10.1371/journal.pbio.1000537

Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69:315–324
Hammer G, Cooper M, Tardieu F, Welch S, Walsh B, van Eeuwijk F, Chapman S, Podlich D (2006) Models for navigating biological complexity in breeding improved crop plants. Trends Plant Sci 11:587–593. doi:10.1016/j.tplants.2006.10.006
Hammer GL, Dong Z, McLean G, Doherty A, Messina CD, Schussler J, Zinselmeier C, Paszkiewicz S, Cooper M (2009) Can changes in canopy and/or root system architecture explain historical maize yield trends in the U.S. corn belt? Crop Sci 49:299–312
Heijnen J (2005) Approximative kinetic formats used in metabolic network modeling. Biotechnol Bioeng 91:534–545. doi:10.1002/bit.20558
Heinrich R, Rapoport TA (1974) A linear steady-state treatment of enzymatic chains. General properties, control and effector strength. Eur J Biochem 42:89–95
Heinrich R, Schuster S (1996) The regulation of cellular systems. Chapman & Hall, New York
Hengge-Aronis R (1999) Interplay of global regulators and cell physiology in the general stress response of Escherichia coli. Curr Opin Microbiol 2:148–152. doi:10.1016/S1369-5274(99)80026-5
Hoogenboom G, White JW (2003) Improving physiological assumptions of simulation models by using gene-based approaches. Agron J 95:82–89
Houle D, Govindaraju DR, Omholt S (2010) Phenomics: the next challenge. Nat Rev Genet 11:855–866
Howard J, Grill SW, Bois JS (2011) Turing's next steps: the mechanochemical basis of morphogenesis. Nat Rev Mol Cell Biol 12:392–398
Hunter PJ, Borg TK (2003) Integration from proteins to organs: the physiome project. Nat Rev Mol Cell Biol 4:237–243. doi:10.1038/nrm1054
Hunter PJ, McCulloch AD, ter Keurs HEDJ (1998) Modelling the mechanical properties of cardiac muscle. Prog Biophys Mol Biol 69:289–331. doi:10.1016/S0079-6107(98)00013-3
Jamshidi N, Palsson B (2008a) Formulating genome-scale kinetic models in the post-genome era. Mol Syst Biol 4:171
Jamshidi N, Palsson B (2008b) Top-down analysis of temporal hierarchy in biochemical reaction networks. PLoS Comp Biol 4, e1000177
Jamshidi N, Palsson B (2010) Mass action stoichiometric simulation models: incorporating kinetics and regulation into stoichiometric models. Biophys J 98:175–185. doi:10.1016/j.bpj.2009.09.064
Jansen RC, Stam P (1994) High resolution of quantitative traits into multiple loci via interval mapping. Genetics 136:1447–1455
Kacser H, Burns JA (1973) The control of flux. Sym Soc Exp Biol 27:65–104
Kauffman KJ, Pajerowski JD, Jamshidi N, Palsson BO, Edwards JS (2002) Description and analysis of metabolic connectivity and dynamics in the human red blood cell. Biophys J 83:646–662. doi:10.1016/S0006-3495(02)75198-9
Kearsey MJ (1998) The principles of QTL analysis (a minimal mathematics approach). J Exp Bot 49:1619–1623
Kennaway R, Coen E, Green A, Bangham A (2011) Generation of diverse biological forms through combinatorial interactions between tissue polarity and growth. PLoS Comp Biol 7, e1002071. doi:10.1371/journal.pcbi.1002071
Keurentjes JJB, Angenent G, Dicke M, Dos Santos V, Molenaar J, van der Putten W, de Ruiter P, Struik PC, Thomma B (2011) Redefining plant systems biology: from cell to ecosystem. Trends Plant Sci 16:183–190. doi:10.1016/j.tplants.2010.12.002
Knapp SJ, Bridges WC Jr, Birkes D (1990) Mapping quantitative trait loci using molecular marker linkage maps. Theor Appl Genet 79:583–592. doi:10.1007/bf00226869
Kromdijk J, Bertin N, Heuvelink E, Molenaar J, de Visser PHB, Marcelis LFM, Struik PC (2014) Crop management impacts the efficiency of quantitative trait loci (QTL) detection and use: case study of fruit load×QTL interactions. J Exp Bot 65:11–22. doi:10.1093/jxb/ert365

Kuchen EE, Fox S, de Reuille PB, Kennaway R, Bensmihen S, Avondo J, Calder GM, Southam P, Robinson S, Bangham A, Coen E (2012) Generation of leaf shape through early patterns of growth and tissue polarity. Science 335:1092–1096. doi:10.1126/science.1214678

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

Laperche A, Devienne-Baret F, Maury O, Le Gouis J, Ney B (2006) A simplified conceptual model of carbon/nitrogen functioning for QTL analysis of winter wheat adaptation to nitrogen deficiency. Theor Appl Genet 113:1131–1146

Li S, Assmann S, Albert R (2006) Predicting essential components of signal transduction networks: a dynamic model of guard cell abscisic acid signaling. PLoS Biol 4, e312. doi:10.1371/journal.pbio.0040312

Ma C-X, Casella G, Wu R (2002) Functional mapping of quantitative trait loci underlying the character process: a theoretical framework. Genetics 161:1751–1762

Ma H-W, Kumar B, Ditges U, Gunzer F, Buer J, Zeng A-P (2004a) An extended transcriptional regulatory network of Escherichia coli and analysis of its hierarchical structure and network motifs. Nucleic Acids Res 32:6643–6649. doi:10.1093/nar/gkh1009

Ma H-W, Zhao X-M, Yuan Y-J, Zeng A-P (2004b) Decomposition of metabolic network into functional modules based on the global connectivity structure of reaction graph. Bioinformatics 20:1870–1876. doi:10.1093/bioinformatics/bth167

Mahadevan R, Schilling CH (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab Eng 5:264–276. doi:10.1016/j.ymben.2003.09.002

Malosetti M, Visser RGF, Celis-Gamboa C, Eeuwijk FA (2006) QTL methodology for response curves on the basis of non-linear mixed models, with an illustration to senescence in potato. Theor Appl Genet 113:288–300

Mammadov J, Aggarwal R, Buyyarapu R, Kumpatla S (2012) SNP markers and their impact on plant breeding. Int J Plant Genomic 2012:11. doi:10.1155/2012/728398

Mao L, Van Hemert JL, Dash S, Ja D (2009) Arabidopsis gene co-expression network and its functional modules. BMC Bioinf 10:346. doi:10.1186/1471-2105-10-346

Martínez-Antonio A, Collado-Vides J (2003) Identifying global regulators in transcriptional regulatory networks in bacteria. Curr Opin Microbiol 6:482–489. doi:10.1016/j.mib.2003.09.002

Martre P, Bertin N, Salon C, Génard M (2011) Modelling the size and composition of fruit, grain and seed by process-based simulation models. New Phytol 191:601–618

Messina CD, Jones JW, Boote KJ, Vallejos CE (2006) A gene-based model to simulate soybean development and yield responses to environment. Crop Sci. doi:10.2135/cropsci2005.04-0372

Messina CD, Hammer GL, Dong Z, Podlich D, Cooper M (2009) Modelling crop improvement in a G×E×M framework via gene-trail-phenotype relationships. In: Sadras VO, Calderini D (eds) Crop physiology: applications for genetic improvement and agronomy. Elsevier, Netherlands, pp 235–265

Miflin B (2000) Crop improvement in the 21st century. J Exp Bot 51:1–8

Monteiro PT, Ropers D, Mateescu R, Freitas AT, de Jong H (2007) Temporal logic patterns for querying dynamic models of cellular interaction networks. Bioinformatics 24:i227–i233

Morris MK, Saez-Rodriguez J, Sorger PK, Lauffenburger DA (2010) Logic-based models for the analysis of cell signaling networks. Biochemistry 49:3216–3224. doi:10.1021/bi902202q

Nägele T, Henkel S, Hörmiller I, Sauter T, Sawodny O, Ederer M, Heyer A (2010) Mathematical modeling of the central carbohydrate metabolism in Arabidopsis reveals a substantial regulatory influence of vacuolar invertase on whole plant carbon metabolism. Plant Physiol 153:260–272

Nakagawa H, Yamagishi J, Miyamoto N, Motoyama M, Yano M, Nemoto K (2005) Flowering response of rice to photoperiod and temperature: a QTL analysis using a phenological model. Theor Appl Genet 110:778–786

Noble D (2002) Modeling the heart–from genes to cells to the whole organ. Science 295:1678–1682. doi:10.1126/science.1069881

Orth JD, Thiele I, Palsson BØ (2010) What is flux balance analysis? Nat Biotechnol 28:245–248. doi:10.1038/nbt.1614

Papin JA, Price ND, Wiback SJ, Fell DA, Palsson BO (2003) Metabolic pathways in the post-genome era. Trends Biochem Sci 28:250–258

Papin JA, Stelling J, Price ND, Klamt S, Schuster S, Palsson BO (2004) Comparison of network-based pathway analysis methods. Trends Biotechnol 22:400–405

Podlich D, Cooper M (1998) QU-GENE: a simulation platform for quantitative analysis of genetic models. Bioinformatics 14:632–653

Prudent M, Bertin N, Génard M, Munos S, Rolland S, Garcia V, Petit J, Baldet P, Rothan C, Causse M (2010) Biological processes and gene categories susceptible to carbon availability in tomato fruit. Plant Cell Environ 33:1186–1204

Prusinkiewicz P, Runions A (2012) Computational models of plant development and form. New Phytol 193:549–569. doi:10.1111/j.1469-8137.2011.04009.x

Quilot B, Génard M, Lescourret F, Kervella J (2005a) Simulating genotypic variations of fruit quality in an advanced peach × Prunus davidiana cross. J Exp Bot 56:3071–3081

Quilot B, Kervella J, Genard M, Lescourret F (2005b) Analysing the genetic control of peach fruit quality through an ecophysiological model combined with a QTL approach. J Exp Bot 56:3083–3092

Radulescu O, Gorban A, Zinovyev A, Lilienbaum A (2008) Robust simplifications of multiscale biochemical networks. BMC Syst Biol 2:86. doi:10.1186/1752-0509-2-86

Reymond M, Muller B, Leonardi A, Charcosset A, Tardieu F (2003) Combining quantitative trait loci analysis and an ecophysiological model to analyze the genetic variability of the responses of maize leaf growth to temperature and water deficit. Plant Physiol 131:664–675. doi:10.1104/pp. 013839

Reymond M, Muller B, Tardieu F (2004) Dealing with the genotype × environment interaction via a modelling approach: a comparison of QTLs of maize leaf length or width with QTLs of model parameters. J Exp Bot 55:2461–2472. doi:10.1093/jxb/erh200

Rios-Estepa R, Lange BM (2007) Experimental and mathematical approaches to modeling plant metabolic networks. Phytochemistry 68:2351–2374. doi:10.1016/j.phytochem.2007.04.021

Schilling CH, Letscher D, Palsson BO (2000) Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. J Theor Biol 203:229–248

Schlitt T, Brazma A (2007) Current approaches to gene regulatory network modelling. BMC Bioinf 8:S9

Schuetz R, Kuepfer L, Uwe S (2007) Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. Mol Syst Biol 3:119

Schuster S, Fell DA, Dandekar T (2000) A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nat Biotechnol 18:326–332

Schuster S, Pfeiffer T, Moldenhauer F, Koch I, Dandekar T (2002) Exploring the pathway structure of metabolism: decomposition into subnetworks and application to mycoplasma pneumoniae. Bioinformatics 18:351–361

Segrè D, Vitkup D, Church GM (2002) Analysis of optimality in natural and perturbed metabolic networks. Proc Natl Acad Sci U S A 99:15112–15117. doi:10.1073/pnas.232349399

Seshasayee A, Fraser G, Babu MM, Luscombe N (2009) Principles of transcriptional regulation and evolution of the metabolic system in E. coli. Genome Res 19:79–91. doi:10.1101/gr.079715.108

Shlomi T, Eisenberg Y, Sharan R, Ruppin E (2007) A genome-scale computational study of the interplay between transcriptional regulation and metabolism. Mol Syst Biol 3:101

Sinclair TR, Purcell LC, Sneller CH (2004) Crop transformation and the challenge to increase yield potential. Trends Plant Sci 9:70–75

Smallbone K, Simeonidis E, Broomhead DS, Kell DB (2007) Something from nothing–bridging the gap between constraint-based and kinetic modelling. FEBS J 274:5576–5585. doi:10.1111/j.1742-4658.2007.06076.x

Snoussi EH (1989) Qualitative dynamics of piecewise-linear differential equations: a discrete mapping approach. Dyn Stab Syst 4:189–207

Southern J, Pitt-Francis J, Whiteley J, Stokeley D, Kobashi H, Nobes R, Kadooka Y, Gavaghan D (2008) Multi-scale computational modelling in biology and physiology. Prog Biophys Mol Bio 96:60–89. doi:10.1016/j.pbiomolbio.2007.07.019

Sridharan GV, Yi M, Hassoun S, Lee K (2012) Metabolic flux-based modularity using shortest retroactive distances. BMC Syst Biol 6:155. doi:10.1186/1752-0509-6-155

Stelling J, Klamt S, Bettenbrock K, Schuster S, Gilles ED (2002) Metabolic network structure determines key aspects of functionality and regulation. Nature 420:190–193. doi:10.1038/nature01166

Stitt M, Sulpice R, Keurentjes JJB (2010) Metabolic networks: how to identify key components in the regulation of metabolism and growth. Plant Physiol 152:428–444. doi:10.1104/pp.109.150821

Sweetlove L, Ratcliffe RG (2011) Flux-balance modeling of plant metabolism. Front Plant Sci 2:1–10. doi:10.3389/fpls.2011.00038

Tanay A, Sharan R, Kupiec M, Shamir R (2004) Revealing modularity and organization in the yeast molecular network by integrated analysis of highly heterogeneous genomewide data. Proc Natl Acad Sci U S A 101:2981–2986. doi:10.1073/pnas.0308661100

Tardieu F (2003) Virtual plants: modelling as a tool for the genomics of tolerance to water deficit. Trends Plant Sci 8:9–14

Tisné S, Serrand Y, Bach L, Gilbault E, Ben Ameur R, Balasse H, Voisin R, Bouchez D, Durand-Tardif M, Guerche P, Chareyron G, Da Rugna J, Camilleri C, Loudet O (2013) Phenoscope: an automated large-scale phenotyping platform offering high spatial homogeneity. Plant J 74:534–544. doi:10.1111/tpj.12131

Tomita M, Hashimoto K, Takahashi K, Shimizu TS, Matsuzaki Y, Miyoshi F, Saito K, Tanida S, Yugi K, Venter JC, Hutchison CA (1999) E-CELL: software environment for whole-cell simulation. Bioinformatics 15:72–84

Uys L, Botha FC, Hofmeyr J-HS, Rohwer JM (2007) Kinetic model of sucrose accumulation in maturing sugarcane culm tissue. Phytochemistry 68:2375–2392

van Eeuwijk F, Bink M, Chenu K, Chapman S (2010) Detection and use of QTL for complex traits in multiple environments. Curr Opin Plant Biol 13:193–205. doi:10.1016/j.pbi.2010.01.001

Visser D, Heijnen J (2003) Dynamic simulation and metabolic re-design of a branched pathway using linlog kinetics. Metab Eng 5:164–176. doi:10.1016/S1096-7176(03)00025-9

Wagner G, Pavlicev M, Cheverud J (2007) The road to modularity. Nat Rev Genet 8:921–931. doi:10.1038/nrg2267

Wang X, Dalkic E, Wu M, Chan C (2008) Gene module level analysis: identification to networks and dynamics. Curr Opin Biotechnol 19:482–491. doi:10.1016/j.copbio.2008.07.011

Welch SM, Roe JL, Dong Z (2003) A genetic neural network model of flowering time control in arabidopsis thaliana. Agron J 95:71–81

Welch SM, Dong Z, Roe JL (2004) Modelling gene networks controlling transition to flowering in arabidopsis. In: Fischer A, Turner N, Angus JF, McIntyre L, Robertson MJ,. Borrell AK, Lloyd D (eds) New directions for a diverse planet: proceedings for the 4th international crop science congress, Brisbane. pp 1–20

Welcker C, Boussuge B, Bencivenni C, Ribaut J-M, Tardieu F (2007) Are source and sink strengths genetically linked in maize plants subjected to water deficit? A QTL study of the responses of leaf growth and of anthesis-silking Interval to water deficit. J Exp Bot 58:339–349

White JW, Hoogenboom G (1996) Simulating effects of genes for physiological traits in a process-oriented crop model. Agron J 88:416–422

White JW, Hoogenboom G (2003) Gene-based approaches to crop simulation: past experiences and future opportunities. Agron J 95:52–64

Wittmann DM, Krumsiek J, Saez-Rodriguez J, Da L, Klamt S, Theis FJ (2009) Transforming Boolean models to continuous models: methodology and application to T-cell receptor signaling. BMC Syst Biol 3:98. doi:10.1186/1752-0509-3-98

Wu W, Zhou Y, Li W, Mao D, Chen Q (2002) Mapping of quantitative trait loci based on growth models. Theor Appl Genet 105:1043–1049

Wu R, Ma C-X, Zhao W, Casella G (2003) Functional mapping for quantitative trait loci governing growth rates: a parametric model. Physiol Genomics 14:241–249. doi:10.1152/physiolgenomics.00013.2003

Yin X, Struik PC (2008) Applying modelling experiences from the past to shape crop systems biology: the need to converge crop physiology and functional genomics. New Phytol 179:629–642

Yin X, Struik PC (2010) Modelling the crop: from system dynamics to systems biology. J Exp Bot 61:2171–2183. doi:10.1093/jxb/erp375

Yin X, Kropff MJ, Stam P (1999) The role of ecophysiological models in QTL analysis: the example of specific leaf area in barley. Heredity 82:415–421

Yin X, Chasalow SC, Dourleijn CJ, Stam P, Kropff MJ (2000a) Coupling estimated effects of QTLs for physiological traits to a crop growth model: predicting yield variation among recombinant inbred lines in barley. Heredity 85:539–549

Yin X, Kropff MJ, Goudriaan J, Stam P (2000b) A model analysis of yield differences among recombinant inbred lines in barley. Agron J 92:114–120

Yin X, Struik PC, Tang J, Qi C, Liu T (2005) Model analysis of flowering phenology in recombinant inbred lines of barley. J Exp Bot 56:959–965

Zeng Z-B (1993) Theoretical basis of separation of multiple linked gene effects on mapping quantitative trait loci. Proc Natl Acad Sci U S A 90:10972–10976

Chapter 2
Simulating Genotype-Phenotype Interaction Using Extended Functional-Structural Plant Models: Approaches, Applications and Potential Pitfalls

Lifeng Xu and Gerhard Buck-Sorlin

Abstract Modelling approaches have increasingly been used as a supplementary tool in understanding the build-up and diversity of crop phenotypes, and their relations with morphogenesis. Among these approaches, Functional-Structural Plant Models (FSPMs) have been developed to simulate complex interactions between plant architecture and physiological processes. In this chapter, we introduce an FSPM of rice that simulates growth and morphology of individual rice plants and of small stands from germination to seed maturity. This model covers selected ecophysiological processes including photosynthesis and sink functions based on a common assimilate pool. We furthermore introduce here for the first time an extension of the rice FSPM with a module for genetics, which constitutes a genotype-phenotype model coupling quantitative genetic information of the phenotypic trait plant height with the morphogenetic rules leading to this composite trait. Lastly, a virtual breeding model is presented: this extended model enables the virtual reproduction of quantitative genetic information and the generation of a new simulated mapping population, in both its phenotypic and genotypic form. Finally, the current pitfalls and problems, and the potential uses of the virtual breeding model are discussed.

L. Xu
College of Computer Science & Technology, Zhejiang University of Technology, Hangzhou 310023, People's Republic of China

G. Buck-Sorlin (✉)
UMR 1345 Institut de Recherche en Horticulture et Semences (IRHS), AGROCAMPUS OUEST, 2 rue André le Nôtre, 49045 Angers Cedex 01, France
e-mail: gerhard.buck-sorlin@agrocampus-ouest.fr

X. Yin, P.C. Struik (eds.), *Crop Systems Biology*,
DOI 10.1007/978-3-319-20562-5_2

2.1 Introduction and Overview: Using Morphogenetic Models in Cereal Breeding

Cereals, such as rice, maize and wheat, constitute the basic staple foods of humankind, and the demand for them is rising with the increase in the human population. Cereal breeding science needs to ensure yield stability and increase in productivity in order to meet this increasing demand: Ensuring the current level of supply with cereals requires a joint effort employing all the technical means presently available. However, in the face of climate change breeding efforts need to be doubled to create new adapted varieties. What is more, some socio-economic problems can probably not be resolved by an increase in cereal production: rising prices due to stock market speculation with food commodities; competing claims for the use of agricultural soils; unequal and unjust distribution of food; or food waste and excessive meat consumption.

Three major tasks for cereal breeding derived from this situation are: first and foremost to create disease-resistant cultivars; second, to increase crop efficiency in terms of product quantity per area as well as quality (nutritional and with respect to disease resistance); and third, to create cultivars that are adapted to abiotic stresses (salt, hydric, nutrient depletion, soil compaction,…) so that crop cultivation can be extended to less favourable soils and climates. It goes without saying that most often these named tasks cannot be tackled separately; therefore, it will be an overarching task of plant breeding to come up with cultivars that are disease-resistant, stress-resistant or stress-tolerant, and efficient in terms of yield.

The contribution of science to solving these problems can be manifold; amongst them is the unravelling of the processes that lead to complex traits such as yield. The interaction of genes (genotypes, genomes…), environmental factors (also in the form of stresses), and crop management with physiological processes at the organ, plant and crop canopy level and its integration over time from germination to maturity determines yield quantity and quality. In strong contrast to animals who adapt to the environment largely by their behaviour, plants exhibit a strong phenotypic plasticity, i.e., when adapting to a certain environment they change their overall shape or architecture, due to the modularity of the plant and its ability to form new phytomers on a regular basis by growth and branching. As this plasticity also affects many agronomically important traits (e.g., culm stability and architecture influencing the infection of the plant with soil-borne fungal spores; or the classical example of introduction of semi-dwarf varieties leading to higher grain yields), it has to be considered in an explicit manner.

Genotype-phenotype models of crops in the form of spatially explicit models of morphogenetic development (commonly in a 3D form), based on the interaction of compartmentalized physiological processes, and coupled with quantitative genetic information, can be used to better understand these systems and ultimately to aid breeding of new crop varieties. Essential elements (objects, rules, and methods) for such a model comprise the description of the phenotype in terms of its constituent organ object types. Another type of elements comprises primary processes and

processes likely to exhibit genetic variation and/or leading to yield traits. Next, the genotype with different degrees of complexity (e.g., arrays of numbers representing genes or "virtual chromosomes" of differing ploidy level or genetic regulatory networks) needs to be described. Finally, a description of processes representing breeding, i.e. selection, recombination, mutation, reproduction, is necessary. Most importantly, such a model should provide the link between the genotype and the phenotype.

Among the objectives of such advanced genotype-phenotype models one could name the production of a large number of recombinations (or a synthetic genotype) and the subsequent testing of the performance of the virtual genotype under a given environment, which is impossible in real breeding; or testing the performance of a given cultivar under different climate conditions (climate change), which is difficult in real breeding. Resulting models should be in the form of multi-scaled and modular systems to facilitate the representation of basic genetic processes and their upscaling to the relevant levels (organ and entire plant). Furthermore, a common ontology for these 3D genotype-phenotype models would have to be developed, and lastly existing data acquisition techniques for this class of models need to be improved. An ultimate objective would be the construction of architectural ideotypes which represent optimal (with respect to the complex trait yield) ranges for individual traits. Such traits may include culm length, tillering intensity, culm leaf number, distribution of leaf blade bases on the culm, leaf blade angle, leaf blade and sheath colour, phyllochron (deviation from distichous arrangement), culm mechanical properties (bending strength, rigidity), grain number per panicle, and panicle architecture (divergence of primary and secondary branches).

Plant models can be used to simulate the genetic characteristics of plants. Conventional crop growth models are now increasingly used in breeding programmes to assist in the design of new plant types (Yin et al. 1999; Hammer et al. 2006). Modelling approaches can be an intuitive and extensible tool to enhance our understanding of complex crop phenotypes, which will ultimately lead to new and improved crop cultivars. Efforts to simulate complex interactions between plant architecture and the physical and biological processes that drive plant development at several temporal and spatial scales have led to the emergence of functional-structural plant models (FSPMs). FSPMs are defined as models that couple a selection of physiological processes that result in an explicit 3D plant structure, often supplied with a mutual feedback between physiology and structure (Buck-Sorlin 2013).

2.2 Functional-Structural Plant Models of Cereals

Understanding plant growth and morphology is of great importance to agronomy. Modelling efforts within the FSPM domain are mostly concerned with the acquisition, transport and use of matter and energy from sources to sinks through pathways dictated by plant architecture, such as light, carbon, water and soil minerals, and how these environmental parameters affect growth and morphology of the resulting

plants (Kitchen and Allaby 2013). FSPMs have been developed for a number of cereal crops, e.g., maize (Fournier and Andrieu 1999), wheat (Fournier et al. 2003), barley (Buck-Sorlin et al. 2008), and rice (Luquet et al. 2006; Xu et al. 2011), see also Fournier et al. (2007). Depending on the application domain, FSPMs integrate different physical and physiological processes and vary in the level of detail for the spatial representation of the plant. Models may consider different hierarchical scales: from individual organs or sets of organs to entire plants.

Application of the FSPM paradigm enables the creation of virtual plants that represent simulations of realistic morphological and physiological development over time. Chenu et al. (2007) simulated leaf development of *Arabidopsis thaliana* as a function of light input. Qu et al. (2012) used an approach based on metamer and root agents to simulate carbon and water acquisitions and expenses in orange trees. Drouet and Pagès (2007) proposed a model for maize during the vegetative period considering carbon and nitrogen dynamics. Clark and Bullock (2007) studied the influence of light competition on plant morphology in a generalized virtual plant.

The above-mentioned studies did not attempt to consider the link between population genetics and individual plant development. Two studies illustrate the efforts to implement genotype-by-environment (G×E) interactions into FSPMs. Buck-Sorlin et al. (2005) developed an FSPM of barley using Relational Growth Grammars (Kniemeyer 2008), in which a graph representation of a genetic regulatory network was used to model the final steps of the biosynthesis of the bioactive form of gibberellic acid – which plays a role in internode extension of barley and other cereals – and its two precursor molecules. Graph topology was also used to represent two virtual chromosomes consisting of seven alleles each, together forming the virtual genotype; genetic processes such as crossing-over and recombination were implemented as growth grammar rules editing and modifying the graph. This permitted the simulation of sexual reproduction, on both the level of the genotype and the phenotype, the latter by mapping the net phenotype effects of each allele to specified morphogenetic rules. Luquet et al. (2012) proposed a rice FSPM with its growth rate being parameterized with different genotype effects.

2.2.1 A New Functional-Structural Plant Model for Rice

In this section we will describe the FSPM of rice, proposed by Xu et al. (2010). This FSPM simulates growth and morphology of the individual rice plant from germination to seed maturity, in combination with selected ecophysiological processes including photosynthesis and sink functions based on a common assimilate pool. The model produces a simple phenotype based on a set of morphogenetic rules describing an "average" development course and final morphology linking yield traits to selected physiological processes. This model has the capability to be extended with a genetic model (see later). Here, we will present some of the main functions of the rice FSPM. The model has several modules, i.e., a photosynthesis module as the source, a growth function and an assimilate partitioning module as

the sink, and a morphology module responsible for the extension of the organs and overall individuals.

In order to model source activity, the photosynthesis model LEAFC3 (Nikolov et al. 1995) was implemented in the XL language (Hemmerling et al. 2008); species-specific parameters for rice were extracted from the literature (Yin et al. 2004; Yin and van Laar 2005; Borjigidai et al. 2006). Input to the photosynthesis model was a weather file (three locations: two in Hangzhou, and one in Hainan, P.R. China) containing daily values of mean temperature, global radiation, and relative humidity. To estimate local light interception and leaf photosynthesis, the ray-tracer-based radiation model of GroIMP, a modelling platform designed for FSPM, was used (Hemmerling et al. 2008). A leaf blade is modelled as a collection of 25 parallelogram objects of different size and orientation. Each leaf has a parameter to store intercepted PAR (Photosynthetically Active Radiation). A method which invokes the LEAFC3 model with input leaf area, temperature and PAR, was used to calculate daily assimilate production per leaf. The output of all leaves was at each step added to a central assimilate pool.

The timing and growth duration of active sinks control the conversion of assimilates to harvestable dry matter, i.e., grain yield. In the FSPM approach of Xu et al. (2010) the orchestration of sink activity is prescribed by growth and development rules and the overall biomass production is an emergent property of the integration of these rules applied over time to simulated structures. In addition, the rate of extension of each organ is described, based on a function for the determinate growth (Yin et al. 2003). This function describes the dynamics of extension and biomass accumulation of organs: the sink strength of a growing organ is first described by its potential growth rate, for which we chose the derivative of the sigmoid growth function proposed by Yin et al. (2003); then the potential growth rate and the realized growth rate at a certain time are calculated (see detailed description in Xu et al. 2010). Once growth of an organ is realized with a rate inferior to or equal to the potential growth rate the central assimilate pool is updated accordingly. Finally, growth respiration is considered in the form of a conversion factor (g glucose g^{-1} newly produced dry matter), which is proportional to the growth rate as described in Goudriaan and van Laar (1994). Likewise, maintenance respiration is computed as a fixed proportion (0.014 g glucose g^{-1} dry matter) of structural biomass. Both terms are subtracted from the central pool at each step.

To simulate vegetative and generative development, a set of growth, developmental and branching rules are repetitively applied to a meristem module and all its ensuing organs, leading to the visible phenotype. The structural framework thus created is used to simulate and analyse the dynamics of assimilate flow as dictated by local (potential) growth rates and assimilate availability in the central pool. Developmental rate, i.e., the rate of formation of new organs per thermal time, is controlled by a parameter (plastochron): each newly formed organ has a state variable p, which is initiated with a value representing the measured duration [d] of the plastochron; at each step p is reduced by one; when it becomes zero the formation of a new phytomer (leaf, internode and lateral tiller bud) is triggered. In addition, the formation of main stem and tiller phytomers is restricted to a maximum phytomer

rank and a maximum tillering order, based on measurements. A leaf is initiated with an initial length and diameter. The meristem is reinitiated at the tip of the shoot, and the rank increased by one; at the same time the plastochron is set to its initial value (as specified by a global variable). Other rules determine bending-up of the culm due to phototropism.

The potential extension and final dimension of organs (leaves, internodes, etc.…) depend upon their rank and age, while the actually achieved dimensions are also a function of sink competition and assimilate availability, as described above.

Leaf dimensions are determined, again using the sigmoid growth function from Yin et al. (2003), calculating dry matter increment as a function of time which is then converted into leaf area.

Once a temperature sum threshold is surpassed, the vegetative meristem is transformed into a generative meristem, which is followed by the formation of grain primordia and grain filling (the latter again described by using the growth function of Yin et al. 2003). As long as the central assimilate pool is not exhausted, the generative meristem then recursively produces grains, thereby diminishing and usually quickly exhausting the assimilate pool.

The central assimilate pool functions as a buffer between sources and sinks. It is replenished by local leaf photosynthesis and diminished by organ growth and maintenance respiration. According to the simulation results by Xu et al. (2010), the dynamics of the central assimilate pool was characterised by three phases. At first, there was an initial decrease due to a strong sink demand by establishment growth accompanied by a lack of source activity due to the unfolding of seedling leaves. This was followed by a strong increase during the mid to late vegetative phase when full photosynthetic capacity is achieved. The final phase was characterized by a sharp drop during the grain filling stage when leaves senesce and sink demand by grains becomes dominant.

This model reproduced plant architecture and morphology of rice plants from the seedling stage to maturity. With the processes of leaf extension, internode elongation, and grain formation (constituting the sink model), the model already exhibits a reasonably faithful appearance at all growth stages, as shown in Fig. 2.1 for the rendered scene of simulated rice morphology with light distribution.

2.3 Coupling the Functional-Structural Plant Model of Rice with Quantitative Genetic Information

As outlined in the introduction the improvement of the genetic pool of rice with the objective to sustainably increase yields has long been the focus of breeders. In order to improve production efficiency as well as product quality, a deeper understanding of the way ecophysiological characteristics functionally contribute to yield, and how genes interact with changing environments to act upon physiological processes, is necessary. With respect to quantitative genetics, recent progress has enabled us to

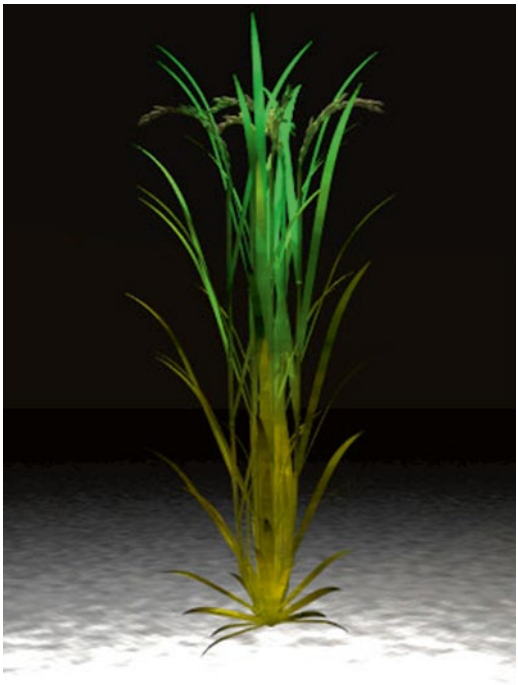


Fig. 2.1 Snapshot of the rendered scene of the simulated rice morphology at the post-flowering stage, using the Twilight renderer of GroIMP to show simulated light distribution

dissect naturally occurring variations and has contributed to our understanding of the genetic control of morphological and physiological traits (Fukuoka et al. 2010). Mapping quantitative trait loci (QTL) for complex traits has become a routine tool in functional genetics and plant breeding research (see Chap. 1 of this book by Baldazzi et al.), and QTL analysis is providing a powerful strategic tool to associate genomic regions with their phenotypic effects in rice (Yamamoto et al. 2010).

Traditional breeding methods are labour-intensive and time consuming which renders the development of new advanced crop cultivars increasingly difficult (Fukuoka et al. 2010). Growth model-based approaches in combination with quantitative genetics can assist in and accelerate traditional breeding (Lecoeur et al. 2011; Uptmoor et al. 2008). Coupling QTL and FSPM allows the spatially explicit modelling of phenotypes and their development, under observed conditions but also under environmental conditions that have not been experimentally tested or of new genotypes that only exist *in silico* (Reymond et al. 2004).

As a first major step towards this goal, Xu et al. (2011) linked our rice FSPM to a quantitative genetic model, thereby employing QTL information to refine model parameters and visualizing the dynamics of the entire phenotype as a result of underlying ecophysiological processes, including the trait for which genetic information is available. This has laid the foundation for future use of the model to design ideotypes.

Xu et al. (2011) combined the ecophysiological model of rice with QTL information on internode extension, which has a direct influence on plant height (plant height being affected by the sum of the lengths of all internodes, erectness of culms, the length and insertion height of the flag leaf blade, and the erectness and size of the panicle). In the extended model yield traits were linked to selected physiological processes and detected QTLs with additive effects and epistatic effects (a gene enhancing or inhibiting the expression of another gene located at a different locus; see also Chap. 1 of this book by Baldazzi et al.). A follow-up study presented an example of a rice model that simulated genotypes superior for plant height (Xu et al. 2012): the basic model represented QTLs as arrays of genotypic values and morphogenesis as rules in relation to other physiological processes. Virtual breeding was conducted as the reproduction of QTLs from parental lines to a mapping population. The rules that specified the genetic processes operating on genotypes as intrinsic properties of each individual determined the phenotype value, e.g., the plant height.

The link between the ecophysiological model and QTL information, and the processes of sexual reproduction with genetic information were constructed via a genetic model, containing an interface responsible for data input (quantitative genetic data), methods that can be invoked to simulate the genetic processes (i.e., crossing-over, recombination, and haploid doubling), the interaction by the user (i.e., selection of individuals as parents for sexual reproduction), and determination of the phenotype coefficient.

The interface was designed as a dialog of the rice FSPM model with the output of QTLNetwork (a software for QTL mapping, Yang et al. 2008). QTL information, i.e., the number of QTLs detected, additional effects, epistasis effects, position information including flanking markers, mean value(s) of the trait(s) in question of the population, can be automatically extracted from the output files of QTLNetwork. Thus, the recombination rates of the loci and total additional phenotype coefficients can be used as model parameters, for the visual representation and for the processes involved in virtual breeding.

The parameters converted from the genetic data and exhibiting variation among individuals of a mapping population were then used in the simulation instead of standard/average model parameters, calculated as

$$y = \mu + G \tag{2.1}$$

where y is the phenotype coefficient that can be used as the parameter in the model, computed as the sum of the population mean (μ) and the total genetic effect (G). The latter consists of genetic main and interactive effects; genotype by environment interaction effect was not considered (Xu et al. 2010). The total genetic effect can then be written as

$$G = \sum_i^n x_i a_i + \sum_i^n \sum_{j=i+1}^n x_i x_j aa_{ij} \tag{2.2}$$

where n denotes the total number of QTLs detected; a_i is the additive effect of the i-th QTL; aa_{ij} is the digenic epistatic effect of additive by additive interaction between i-th and j-th QTL; x_i and x_j refer to the genotypes of the two QTLs, respectively.

Generally speaking, y is a prediction of the observed phenotype (e.g., stem length, tiller number, grain number, grain weight, etc.), computed from the mixed linear model, for a given genotype and environment (Xu et al. 2011).

Take the trait plant height as an example: first, an algorithm implementing the derivative of the beta growth function was employed. Using this function, plant height was then calculated from the population mean value and predicted genetic effects were used to tune the growth curve of the entire stem. More specifically, the instant growth rates and final dimensions (final length, analogous to w_{max}, used in the derivative of the aforementioned sigmoid growth function) of each internode were used to reproduce the length dynamics of the stem. For simplicity, a fixed length distribution and internode number were assumed.

The mechanisms of sexual reproduction considering representations of a QTL genotype are based on the implementation of "biomorphs" (Kniemeyer et al. 2003) and the "virtual breeding" model *BarleyBreeder* (Buck-Sorlin et al. 2005, 2007). Here, we will discuss a number of rules in the rice FSPM, with which the sexual reproduction of individuals with marker genotypes and QTLs was implemented.

Consider the following rule:

```
Axiom ==>Population Genotype [ch1:Chromo ch2:Chromo];
```

In this initiation rule (using the start word *Axiom*), each chromosome (*Chromo*) is declared separately as a branch (designated by opening and closing square brackets "*[*" and "*]*") of *Genotype*. In turn, *Genotype* contains a variable *qtl*, which is an array of the QTL genotype containing the value 1 or −1 (identified for *Q* or *q*, respectively, as the possible allele) in each locus.

QTLs are regions on a chromosome, characterized by a pair of (molecular) markers, which are statistically associated with a trait. This has been imitated in our model. Thus, the effects for each QTL is an array *Genotype*. *QTL_Effect*[i] from the node *Genotype*:

```
Genotype.setQTLeffect(ch1.qtl, ch2.qtl)
```

where the method *setQTLeffect()* is a method of *Genotype* which computes the phenotypic effect of a given QTL, *ch1* and *ch2* represent the two chromosomes (object *Chromo*), while *ch1.qtl* and *ch2.qtl* are two arrays indicating the QTL genotype.

A typical run of the model starts with two default individuals considered as parents (P1, P2) of the doubled haploid (DH) population (Fig. 2.2). These parental individuals represent two instances of the same set of developmental rules yet linked to two different genotypes; at this stage the development of the parental individuals can be simulated to visualize their final morphological phenotypes. Initial variables (Table 2.1) are read from the external files (QTLNetwork output files). The QTLs are considered as markers, and added into the linkage map with the genetic distance information to form the original genotype specified in genotype initialization as Relational Growth Grammar (RGG) rules:

Fig. 2.2 Simulated final morphology of the two parental lines (P1:'IR64' (*left*), and P2:'Azucena' (*right*)) of the DH population, as the initial two individuals of the rice model

Table 2.1 Predicted genetic effects (G) for P1, P2 on final plant height (in cm) of rice. G_G General genetic effect: G_1, G_2, and G_3 are the total genetic effects in three environments (spots), respectively. The mean value of plant height of the population (μ) is 104.11 cm

	G_G	G_1	G_2	G_3
P1	−25.27	−23.09	−27.19	−25.27
P2	18.00	15.81	19.92	18.00

```
Germ -GEN_EDGE->gen:Genotype [ch1:Chromo ch2:Chromo]
  ::> {
        ch1.setMarker(marker_genotype_input[id]);
        ch2.setMarker(marker_genotype_input [id]);
        ch1.getGenotype();
        ch2.getGenotype();
        gen.setQTLeffect(ch1.chromo_genotype, ch2. chromo_genotype);
      }
```

Here, marker genotypes (*marker_genotype_input[ ]*) are initialized for the two chromosomes (*Chromo*, represented by *ch1* and *ch2*). Then the QTL genotype (*chromo_genotype*) for each allele is extracted from the marker array using the function *getGenotype()*. In this way the QTL effect for an individual (identified by the integer *id*) can be calculated.

The reproduction process can take place interactively at any developmental stage. An F1 generation of virtual rice is then simulated after chromosome alignment, crossing-over and recombination have taken place, with the markers as well as QTLs having been reinitiated. The recombination rates between markers are calculated using the position information with the Haldane map function (Haldane and Waddington 1931; Zeng 2000). The markers of the chosen allele are defined as

the input of the gametal allele using the function *setGenotype()*. The simplified RGG rules for chromosome crossing-over and chromosome recombination are:

```
c11.setGenotype(crossOver(c11.chromo_genotype, c12.chromo_
genotype));
```

where the markers on *c11* (one of the chromosomes) are defined as the product of crossing-over between *c11* and *c12*; the same applies to the other alleles. The function *crossOver()* simulates the crossing-over process, in which the rate of crossing-over of marker alleles is a function of the recombination frequency, the latter being calculated with the Haldane map function (Haldane and Waddington 1931; Zeng 2000):

$$r = \frac{1}{2}(1 - e^{-2x}) \tag{2.3}$$

where x is the genetic distance between markers in a chromosome with the unit Morgan (M, 1 M = 100 centi-Morgan).

Using the *RiceBreeder* model, various populations can be simulated. The DH population is derived from the F1 generation after production of recombined haploids and reduplication of alleles (as shown in Fig. 2.3). The simulated phenotype value as well as the marker genotype for each individual can be obtained after completion of growth. The simulated DH population can be then used as the mapping population in the QTL mapping process, aiming at the validation of the genetic model within the *RiceBreeder*.

2.4 *RiceBreeder*: A Tool for Visual Virtual Breeding

Virtual breeding denotes, essentially, a simplified imitation of real breeding. It should involve the following:

- a population of virtual plant individuals;
- model representation of a phenotype: i.e., 3D morphology, basic physiological processes, transport and environmental sensitivity in an FSPM, or physiological processes plus environmental factors in a crop model;
- model representation of a genotype: sets of variables, i.e., genes and QTLs, with values representing alleles or a set of gene regulatory networks;
- model provisions to link genotype with phenotype and to simulate breeding mechanisms (selection, recombination, mutation, reproduction, etc.).

Virtual breeding models can systematically produce a large number of recombinant genotypes, and predict the performance of the virtual phenotype tested under a given environment. This is almost impossible in real breeding practice. Besides, virtual breeding models can also be used to test the performance of a given cultivar

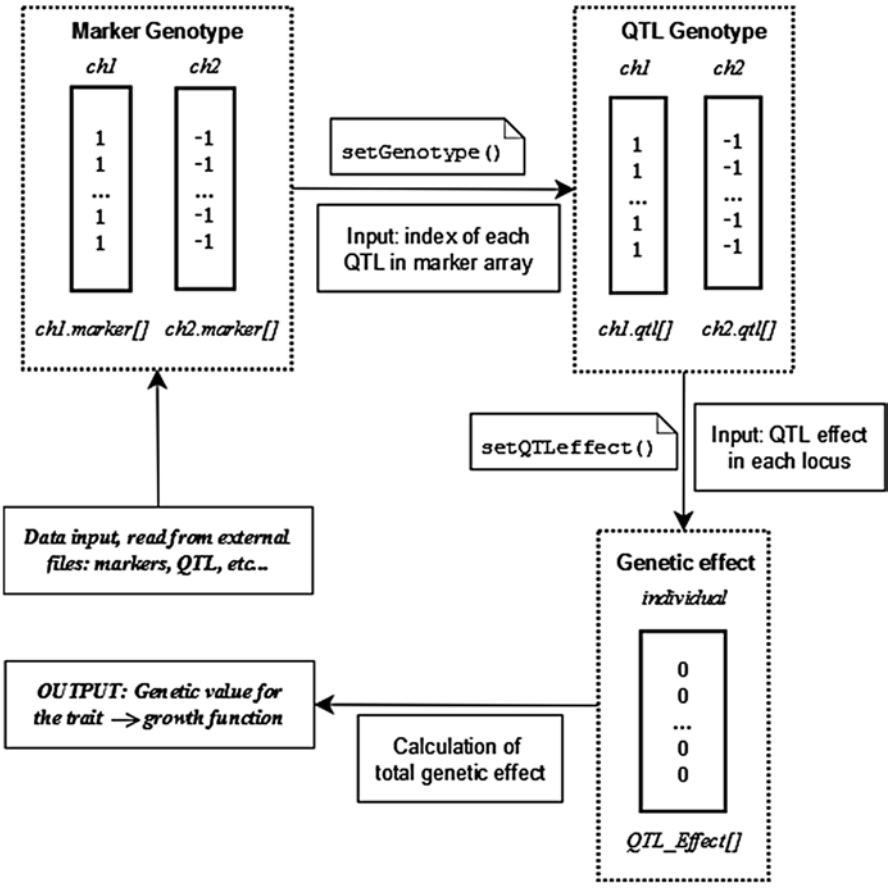


Fig. 2.3 Workflow of the genetic model within the rice FSPM after reading genetic data from the result files of QTL mapping. Taking one individual from the F1 generation as an example, the modules represent marker arrays, QTL genotype and the total genetic effect, respectively. The latter two will not be all zero if the alleles are homozygous

under different climate conditions. In contrast, in real breeding the climatic conditions cannot be fully controlled and thus it will be difficult to get suitable results, not to mention the increased investment in effort, time, and labour necessary in real breeding practice.

Then which physiological processes should be modelled? To answer this question, one should be clear about the basic physiological processes involved in primary production, the processes that exhibit genetic variability and/or lead to yield-related traits. These are the processes that need to be considered. Regarding the number of additional processes, a choice has to be made that balances the risk of over-parameterization with the risk of over-simplification. As a guide, processes should be added until model calibration becomes cumbersome or is no longer possible.

Furthermore, genes or groups of genes indicated by a QTL do not directly control phenotypic traits but underlying processes that have control over the formation of those traits. The biomass quantity, and the quality of harvestable yield depend on organ morphology and plant architecture, i.e., on the structure of the entire crop plant which carries the harvestable parts. Furthermore, the build-up of this structure is influenced by the environment, leading to spatial heterogeneity of physiological processes that are exposed to different sets of microclimates according to the position and exposition of the organ within the canopy. This is an aspect in favour of the use of a spatially explicit FSPM in combination with genotype information. Take the light microclimate in the cut-rose production system as an example (Buck-Sorlin et al. 2011): the amount of light intercepted not only depends on the intensity of the light source, but also on the temporal and spatial dynamics of the architecture of the very heterogeneous canopy, which is influenced by the constant harvest of flowering stems, and that in turn triggers basal bud break and the formation of new flowering shoots. These management factors thus have impact on assimilate production, and together with the other processes, such as assimilate partitioning and organ extension, influence the quality and final amount of flowering stems that can be harvested.

As described in the last section, our FSPM of rice has been extended from the morphological model to a model integrated with QTL information as well as genetic reproduction processes, as the newly formed model *RiceBreeder*. The link between phenotype, presented as the simulated rice plant, and the QTL genotype was implemented via a data interface between the rice FSPM and QTLNetwork. In the example of plant height and grain yield of rice (Xu et al. 2011), the element that connects genetic information with the phenotype trait is the growth function of Yin et al. (2003). The derivative of the growth function determines the extension rate of each growth unit (organ, such as internode). Simulations were run with the parameters calculated from the input QTL data, thereby reproducing variation among individuals.

As the growth rates were controlled by the growth function (w_{max} calculated from the genotype and QTL effects) and its input parameters, some of which are genetically related, the growth curves of the same internode in the four individuals (two parental lines and two superior lines of the rice DH population) are distinct from each other: the positive superior line had the highest growth rate throughout the entire growth period of the internode, the negative superior genotype had the lowest, while the growth rates of the two parental lines were intermediate. The dynamics of the stem length for the four individuals simulated are accordingly different from each other (see the comparison of the dynamics of growth rates of one internode and stem length extension through all growth periods of the four individuals in Xu et al. 2012).

Simulations of virtual reproduction comprise 390 ‘observations’ of simulated phenotype data from 130 lines in the DH population in three environments. Plant height of each line and the marker genotype were recorded after completion of growth. Using the simulated DH population for QTL mapping, with the phenotype value and marker genotype computed in the mapping software QTLNetwork

Table 2.2 Comparison of the QTL mapping results from the simulated DH population with the initial settings for parental lines in the beginning of the simulation (as the observed data, 'Obs.'). QTL position, additive effects (A), as well as QTL × environment effects (AE) from the simulated population (Est.) and initial setting for the markers (Obs.) were compared, and bias of estimations (Bias) was calculated accordingly [i.e., Bias = (Est. – Obs)/Obs.]

	Position		A		AE1		AE2		AE3	
QTL	Est. (Bias)	Obs.	Est. (Bias)	Obs.	Est. (Bias)	Obs.	Est. (Bias)	Bias	Est.	Obs.
1	214.2 (0.05)	204.6	−15.56 (−0.04)	−16.28	0	0	0	0	0	0
2	158.8 (0.05)	151.8	6.44 (0.06)	6.07	0	0	0	0	0	0
3	315.2 (0.06)	297.2	−5.43 (−0.05)	−5.70	0	0	0	0	0	0
4	141.5 (0.01)	140.8	−5.34 (−0.07)	−5.72	2.06 (0.06)	2.18	−1.85 (−0.04)	−1.92	0	0
5	2 (–)	–	−0.22 (–)	–	0	–	0	–	0	–

The single '–' in QTL 5 indicates false positive estimation (there was no such QTL in the parental lines)

(Version 2.2), results were derived and compared with the initial settings (Table 2.2). All the estimations of the main effect QTLs from the simulated DH population were quite close to the true value (bias not higher than 0.07), except for the estimation of the last QTL, which was a false positive locus of the main additive effect. Three false positive estimations of an epistasis effect were detected, while the last estimation of the AA effect had a low bias, as well as one of the two position estimations.

The *RiceBreeder* model reproduces plant architecture, morphology and the QTLs for plant height of rice from germination to maturity. A view on the final morphology of the two parental lines with different homozygous genotypes from the mapping population is already shown in Fig. 2.2, while Fig. 2.4 shows the representation of the five tallest and five shortest rice individuals from the simulated DH population at the final stage. The genetic values for those individuals were estimated by the equations used in QTLNetwork (Version 2.2). Implementation of other morphological parameters and physiological processes was based on those methodologies described in the last section.

The use of the *RiceBreeder* as an educational tool is another interesting application. This was in fact realized between 2012 and 2014 in the frame of a Master level course held for students of agronomical and horticultural engineering at the University of Angers (France) and the National School for Horticulture (Agrocampus Ouest). Students were given an introduction to FSPM, to the *RiceBreeder* model and the functioning of the QTLNetwork software. Then groups of three students were asked to use the *RiceBreeder* and the QTLNetwork software to conduct cycles of virtual breeding and QTL analysis over up to six generations according to self-defined selection criteria and to document their results. The somewhat surprising

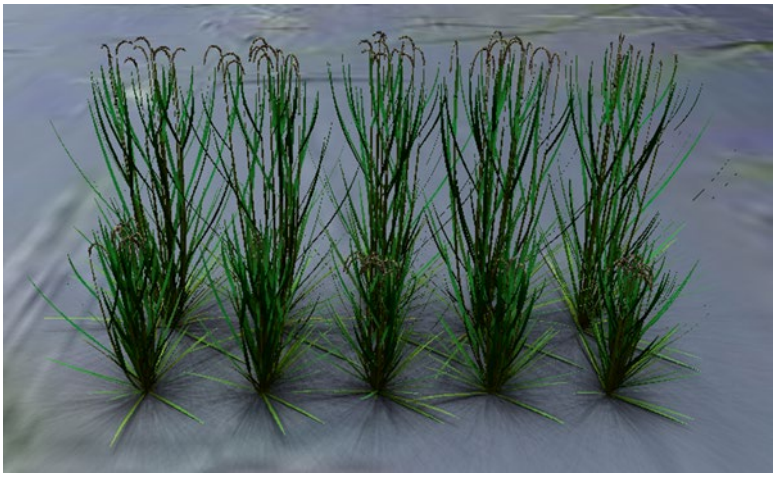


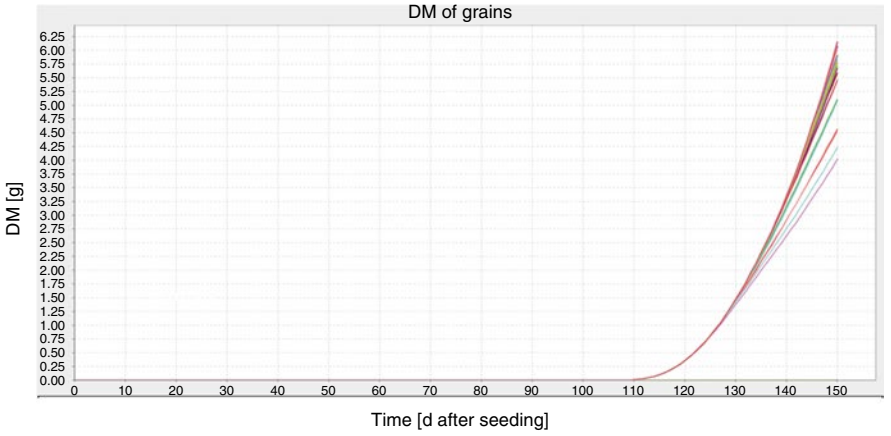
Fig. 2.4 Simulated final morphology of ten lines (*side view*): five highest and five lowest lines selected from the simulated DH population (derived from 'IR64' × 'Azucena') with different QTL genotype for plant height

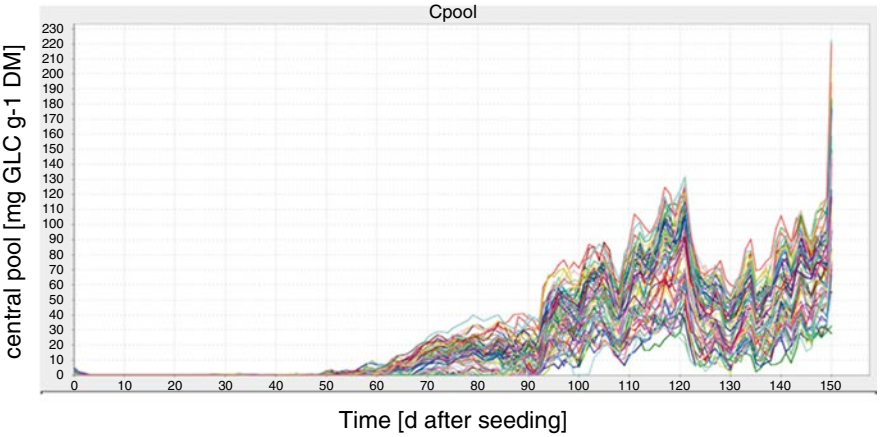
outcome of this exercise was that the students adopted the model rather rapidly and enthusiastically and that they developed a great creativity with respect to the selection criteria, while staying perfectly aware of the limitations of the present approach (see a next section of this chapter). Two general selection strategies emerged: the first strategy consisted in choosing both parents with maximum or minimum values for a certain trait (e.g., the two tallest individuals of a population), whereas the second strategy was to choose the two individuals with the lowest or highest value of the trait considered (e.g., the smallest and largest). Often, a certain trait was monitored while a completely different trait was selected for. QTLs were found to crop up and disappear again from one generation to the next. Figure 2.5 illustrates some results.

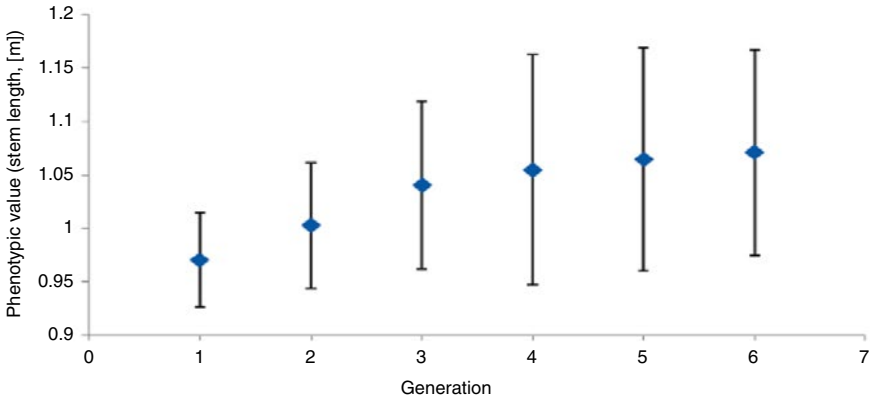
2.5 Current Pitfalls and Problems

The complexity of such an FSPM is not trivial and would be the first potential pitfall/problem that modellers need to solve. More importantly, the development of an extensive simulation tool may not be without problems if inadequate care is put into the development process, or if there is ambiguity in its function, as this may make the tool difficult to communicate or reproduce (Kitchen and Allaby 2013).

In contrast to the real complexity of plant physiological processes, their genetic regulation is often oversimplified in the model. This leads to the second potential pitfall, the relatively inaccurate or over-simplified representation of some processes, especially those describing the functioning of genes and the formation of pheno-


DM of grains
DM [g]
Time [d after seeding]


Cpool
central pool [mg GLC g-1 DM]
Time [d after seeding]


Phenotypic value (stem length, [m])
Generation

typic traits. The number of measurements to establish parameters of morphogenesis and geometry is often too low, not least due to the rather large number of model parameters that have to be considered.

The next problem concerns what could be called "background noise" and is a direct consequence of the above-mentioned simplification. It implies that a vast number of genes in the genome, of physiological processes, and, to some extent, of environmental factors are not considered in any given model yet. In statistics this quantity is usually handled as an error term and can be partitioned into different errors, with respect to their origin (genetic, environment, G×E interaction), and most advanced QTL analysis software takes care of this. Likewise in the current model the error term is then introduced as a stochastic quantity added to the parameter value. However, even if this works fine from a statistical point of view, we should be aware that this does not explain the interaction of the background with the "foreground" (the processes and parameters that are considered in the model). As a consequence, a model which has been calibrated for one genetic background cannot simply be applied to another genetic background without proper recalibration.

Another issue is the optimal size of the simulated population. The threshold size of any mapping population, depending on the level of detail of the mapping objective, should not be less than 100 individuals, while populations used for fine-mapping of traits often exceed 1,000 individuals. Populations of this size cannot be handled for modelling if at the same time the explicit morphological phenotype is to be visualized. Our current version of the *RiceBreeder* handles populations consisting of 60 individuals on a normal laptop computer with two cores without problems, and can visualize the simultaneous development of the entire population within about 15 min. However, the computational power needed for larger populations of simulated plants within FSPM studies quickly exceeds that of a laptop computer; therefore, these simulations would require high-performance computational facilities.

Finally, complex FSPMs of an entire mapping population require a lot of data, for both model calibration and validation, and often the traits needed as model parameters are difficult to establish. High-throughput phenotyping methods are currently being evaluated to automatically measure morphometric traits but they are too coarse and generally only suitable for phenotyping of plant level traits such as plant height or leaf area distribution. With current high-throughput phenotyping

Fig. 2.5 Outcome of a virtual breeding exercise conducted by Master of Science students agronomy. Simulated temporal dynamics of several traits for 60 individuals of a virtual mapping population. *Upper panel*: Trait 'dry mass of grains' (mg), in the second generation. Selection was carried out at day 150 (final step) by choosing individuals with a high grain dry mass. One QTL was found on chromosome 1 with a heritability of 20.7% ($P = 0.006$). Note that the selection was continued until generation 5, yet that the effect of the QTL disappeared afterwards. *Middle panel*: Trait 'central carbon pool', in the fifth generation. Selection was carried out at day 50 by choosing individuals with a low carbon pool. A QTL was found on chromosome 11 with a heritability of 21.6% ($P = 0.00427$). *Lower panel*: Mean value (points) and standard deviation (bars) of the trait 'stem length' over five simulated cycles of virtual breeding. Selection was carried out at day 150 by choosing the two tallest individuals

platforms, multiple linear regression models are conventionally used to generate traits that are difficult to measure, from topological or morphometric traits that are far easier to measure.

2.6 How Can Virtual Breeding Become a Useful Tool in the Near Future?

The potential use of virtual breeding as a tool should in our opinion mirror the three application domains of FSPM: use in education, use in scientific research, and visual decision support.

Regarding the first use it can be stated from experience that the visual explanation, in an integrated (cross-domains) dynamic 3D model, of cereal development, its eco-physiology and genetics, greatly helps to kindle an initial interest in the student which can then be fomented by more specific exposure to subject-bound knowledge. FSPM is an excellent tool to illustrate both phenotypic plasticity and the morphological development of mutants. Accordingly, virtual breeding can be employed in the context of an FSPM population model to teach the genetic basis and practice of selection strategies in both an educational and time-saving manner.

As far as the use as a research tool is concerned, FSPM helps to integrate and visualize metadata, from data-mining of scientific articles and unpublished data across disciplines, and own measured data. Kniemeyer and Buck-Sorlin (unpublished) coupled an abstract hormonal network (consisting of genes, transcription factors and hormones) from the literature (using parameter values from the enzyme databases Brenda and Expasy) with morphogenetic models of *Arabidopsis* and barley to simulate overexpression or loss-of-function mutants. Of course, such models cannot be numerically validated but can nevertheless serve to check for plausibility of a number of contrasting hypotheses as has also been confirmed (e.g., Luquet et al. 2006, 2007).

The use of virtual breeding in decision-support has not yet been achieved. In order to arrive at this point within the next 5–10 years, a number of measures need to be taken to improve the applicability and reliability of this approach:

- the creation of virtual plants that are dynamic, visual, object-oriented, and flexible tools that can be linked to (meta-)databases,
- the modelling of G×E×P interactions at lower levels using networks, where 'P' represents the physiological processes on which the genotype and the environment exert their effects,
- the modelling of complex interaction networks among gene loci, along with a better handling of QTL data,
- an improved description of physiological and morphogenetic processes, leading to truly multi-scaled modelling at organ and plant level, as well as interfaces to other decision-support tools,
- and, the use of genetic algorithms in plant type design.

Current developmental state of the RiceBreeder model, and possible extensions
The current version of the *RiceBreeder* has been outlined in the previous sections. As with any extensive modelling project the wish list for future extensions becomes longer with time, and priorities have to be set in order to maintain model integrity and functionality. Amongst the more urgent things to be considered in future versions are:

- the implementation of more physiological functions, especially with respect to growth hormones such as auxins, gibberellic acids and cytokinins, in the form of simple regulatory networks derived from the literature;
- the explicit consideration of root architecture and soil composition. We have implemented a general, 3D root architecture and soil model, which we intend to parameterize for rice and then couple with the current rice FSPM.
- the implementation of an automatic sensitivity analysis, by which the parameters and model modules necessary for a certain application are filtered out of the complete set of available modules, thereby streamlining the model. The approach by Cournède et al. (2013), which was especially developed for FSPM, will be used.

Finally, the existing software tools will be further extended with the help of information scientists and bioinformaticians. This concerns largely the software GroIMP, in which the following extensions are foreseen:

- the implementation of a functionality for explicit up- and down-scaling,
- an interface to validation tools (see above),
- the extension of modelling techniques, i.e., the development of modular "bricks" to facilitate the representation of basic genetic processes and their up-scaling to the levels relevant for FSPM (organ and plant).

These above-mentioned extensions have to go along with the development of a general framework (or "ontology") for 3D genotype-phenotype models, which does not yet exist, as well as improved data acquisition techniques for genotype-phenotype models. This concerns both high-throughput phenotyping techniques and statistical methods to reduce sample size.

In the future cereal breeders could use virtual plants to validate and monitor data from high-throughput phenotyping experiments and to show up knowledge gaps. Furthermore, the functionality of virtual plants (comprising basic physiological process descriptions encapsulated in organ-level units and fully modularized) would allow testing the performance of an ideotype *in silico*.

References

Borjigidai A, Hikosaka K, Hirose T et al. (2006) Seasonal changes in temperature dependence of photosynthetic rate in rice under a free-air CO_2 enrichment. Ann Bot 97:549–557

Buck-Sorlin GH (2013) Functional-structural plant modelling. In: Dubitzky W, Wolkenhauer O, Cho K, Yokota H (eds) Encyclopedia of systems biology. Springer, New York, pp 778–781

Buck-Sorlin GH, Kniemeyer O, Kurth W (2005) Barley morphology, genetics and hormonal regulation of internode elongation modelled by a relational growth grammar. New Phytol 166:859–867

Buck-Sorlin GH, Kniemeyer O, Kurth W (2007) A grammar-based model of barley including virtual breeding, genetic control and a hormonal metabolic network. In: Vos J, Marcelis LFM, de Visser PHB, Struik PC, Evers JB (eds) Functional-structural plant modelling in crop production. Springer, The Netherlands, pp 243–252

Buck-Sorlin GH, Hemmerling R, Kniemeyer O et al. (2008) A rule-based model of barley morphogenesis, with special respect to shading and gibberellic acid signal transduction. Ann Bot 101:1109–1123

Buck-Sorlin GH, de Visser PHB, Henke M et al. (2011) Towards a functional-structural plant model of cut-rose – simulation of light environment, light absorption, photosynthesis and interferences with the plant structure. Ann Bot 108:1121–1134

Chenu K, Franck N, Lecoeur J (2007) Simulations of virtual plants reveal a role for SERRATE in the response of leaf development to light in Arabidopsis thaliana. New Phytol 175:472–481

Clark B, Bullock S (2007) Shedding light on plant competition: modelling the influence of plant morphology on light capture (and vice versa). J Theor Biol 244:208–217

Cournède PH, Chen Y, Wu QL et al. (2013) Development and evaluation of plant growth models: methodology and implementation in the Pygmalion platform. Math Model Nat Phenom 8(4):112–130

Drouet JL, Pagès L (2007) GRAAL-CN: a model of growth, architecture and allocation for carbon and nitrogen dynamics within whole plants formalised at the organ level. Ecol Model 206:231–249

Fournier C, Andrieu B (1999) ADEL-maize: an L-system based model for the integration of growth processes from the organ to the canopy. Application to regulation of morphogenesis by light availability. Agronomie 19:313–327

Fournier C, Andrieu B, Ljutovac S et al. (2003) Adel-wheat: a 3D architectural model of wheat development. In: Hu B, Jaeger M (eds) Plant growth modeling and applications. Tsinghua University Press/Springer, Beijing, pp 54–66

Fournier C, Andrieu B, Buck-Sorlin GH et al. (2007) Functional-structural modelling of gramineae. In: Vos J, Marcelis LFM, de Visser PHB, Struik PC, Evers JB (eds) Functional-structural plant modelling in crop production. Springer, Dordrecht, pp 175–186

Fukuoka S, Nonoue Y, Yano M (2010) Germplasm enhancement by developing advanced plant materials from diverse rice accessions. Breed Sci 60:509–517

Goudriaan J, van Laar HH (1994) Modelling potential crop growth processes. Kluwer, Dordrecht

Haldane JBS, Waddington CH (1931) Inbreeding and linkage. Genetics 16:357–374

Hammer GL, Cooper M, Tardieu F et al. (2006) Models for navigating biological complexity in breeding improved crop plants. Trends Plant Sci 11:587–593

Hemmerling R, Kniemeyer O, Lanwert D et al. (2008) The rule-based language XL and the modelling environment GroIMP illustrated with simulated tree competition. Funct Plant Biol 35:739–750

Kitchen JL, Allaby RG (2013) Systems modeling at multiple levels of regulation: linking systems and genetic networks to spatially explicit plant populations. Plant 2:16–49

Kniemeyer O (2008) Design and implementation of a graph grammar based language for functional-structural plant modelling. Doctoral dissertation, University of Technology at Cottbus

Kniemeyer O, Buck-Sorlin GH, Kurth W (2003) Representation of genotype and phenotype in a coherent framework based on extended L-systems. Lect Notes Artif Intel 2801:625–634

Lecoeur J, Roire-Lassus R, Christophe A et al. (2011) Quantifying physiological determinants of genetic variation for yield potential in sunflower. SUNFLO: a model-based analysis. Funct Plant Biol 38:246–259

Luquet D, Dingkuhn M, Kim H et al. (2006) EcoMeristem, a model of morphogenesis and competition among sinks in rice. 1. Concept, validation and sensitivity analysis. Funct Plant Biol 33:309–323

Luquet D, Song YH, Elbelt S et al. (2007) Model-assisted physiological analysis of *phyllo*, a rice architectural mutant. Funct Plant Biol 34:11–23

Luquet D, Soulié JC, Rebolledo MC (2012) Developmental dynamics and early growth vigour in rice. 2. Modelling genetic diversity using ecomeristem. J Agron Crop Sci 198:385–398

Nikolov NT, Massman WJ, Schoettle AW (1995) Coupling biochemical and biophysical processes at the leaf level: an equilibrium photosynthesis model for leaves of C3 plants. Ecol Model 80:205–235

Qu H, Wang Y, Cai L (2012) Orange tree simulation under heterogeneous environment using agent-based model ORASIM. Simul Model Pract Theor 23:19–35

Reymond M, Muller B, Tardieu F (2004) Dealing with the genotype × environment interaction via a modelling approach: a comparison of QTLs of maize leaf length or width with QTLs of model parameters. J Exp Bot 55:2461–2472

Uptmoor R, Schrag T, Stuetzel H et al. (2008) Crop model based QTL analysis across environments and QTL based estimation of time to floral induction and flowering in *Brassica oleracea*. Mol Breed 21:205–216

Xu L, Henke M, Zhu J et al. (2010) A rule-based functional-structural model of rice considering source and sink functions. In: Li B, Jaeger M, Guo Y (eds) Plant growth modelling, simulation, visualisation and applications. Proceedings–PMA09. IEEE Computer Society, Los Alamitos, pp 245–252

Xu L, Henke M, Zhu J et al. (2011) A functional–structural model of rice linking quantitative genetic information with morphological development and physiological processes. Ann Bot 107:817–828

Xu L, Ding W, Zhu J et al. (2012) Simulating superior genotypes for plant height based on QTLs: towards virtual breeding of rice. In: Kang M, Dumont Y, Guo Y (eds) 2012 IEEE 4th international symposium on plant growth modeling, simulation, visualization and applications (PMA12), Shanghai, pp 447–454

Yamamoto T, Nagasaki H, Yonemaru J et al. (2010) Fine definition of the pedigree haplotypes of closely related rice cultivars by means of genome-wide discovery of single-nucleotide polymorphisms. BMC Genomics 11:267. doi:10.1186/1471-2164-11-267

Yang J, Hu C, Hu H et al. (2008) QTLNetwork: mapping and visualizing genetic architecture of complex traits in experimental populations. Bioinformatics 24:721–723. doi:10.1093/bioinformatics/btm494

Yin X, van Laar HH (2005) Crop systems dynamics: an ecophysiological simulation model for genotype-by-environment interactions. Wageningen Academic, Wageningen

Yin X, Kropff MJ, Stam P (1999) The role of ecophysiological models in QTL analysis: the example of specific leaf area in barley. Heredity 82:415–421

Yin X, Goudriaan J, Lantinga EA et al. (2003) A flexible sigmoid function of determinate growth. Ann Bot 91:361–371 (with erratum in Ann Bot 91:753, 2003)

Yin X, Van Oijen M, Schapendonk AHCM (2004) Extension of a biochemical model for the generalized stoichiometry of electron transport limited C_3 photosynthesis. Plant Cell Environ 27(10):1211–1222

Zeng Z (2000) Statistical methods for mapping quantitative trait loci, program in statistical genetics. North Carolina State University, North Carolina

Chapter 3
Modelling of Genotype by Environment Interaction and Prediction of Complex Traits across Multiple Environments as a Synthesis of Crop Growth Modelling, Genetics and Statistics

Daniela Bustos-Korts, Marcos Malosetti, Scott Chapman, and Fred van Eeuwijk

Abstract Selection processes in plant breeding depend critically on the quality of phenotype predictions. The phenotype is classically predicted as a function of genotypic and environmental information. Models for phenotype prediction contain a mixture of statistical, genetic and physiological elements. In this chapter, we discuss prediction from linear mixed models (LMMs), with an emphasis on statistics, and prediction from crop growth models (CGMs), with an emphasis on physiology. Three modalities of prediction are distinguished: predictions for new genotypes under known environmental conditions, predictions for known genotypes under new environmental conditions, and predictions for new genotypes under new environmental conditions.

For LMMs, the genotypic input information includes molecular marker variation, while the environmental input can consist of meteorological, soil and management variables. However, integrated types of environmental characterizations obtained from CGMs can also serve as environmental covariable in LMMs. LMMs consist of a fixed part, corresponding to the mean for a particular genotype in a particular environment, and a random part defined by genotypic and environmental

D. Bustos-Korts
Biometris, Wageningen University & Research Centre,
P.O. Box 16, 6700 AC Wageningen, The Netherlands

C.T. de Wit Graduate School for Production Ecology & Resource Conservation (PE&RC), Wageningen, The Netherlands

M. Malosetti • F. van Eeuwijk (✉)
Biometris, Wageningen University & Research Centre,
P.O. Box 16, 6700 AC Wageningen, The Netherlands
e-mail: fred.vaneeuwijk@wur.nl

S. Chapman
CSIRO Plant Industry and Climate Adaptation Flagship, Queensland Bioscience Precinct,
306 Carmody Road, St Lucia, QLD 4067, Australia

X. Yin, P.C. Struik (eds.), *Crop Systems Biology*,
DOI 10.1007/978-3-319-20562-5_3

variances and correlations. For prediction via the fixed part, genotypic and/or environmental covariables are required as in classical regression. For predictions via the random part, correlations need to be estimated between observed and new genotypes, between observed and new environments, or both. These correlations can be based on similarities calculated from genotypic and environmental covariables. A simple type of covariable assigns genotypes to sub-populations and environments to regions. Such groupings can improve phenotype prediction.

For a second type of phenotype prediction, we consider CGMs. CGMs predict a target phenotype as a non-linear function of underlying intermediate phenotypes. The intermediate phenotypes are outcomes of functions defined on genotype dependent CGM parameters and classical environmental descriptors. While the intermediate phenotypes may still show some genotype by environment interaction, the genotype dependent CGM parameters should be consistent across environmental conditions. The CGM parameters are regressed on molecular marker information to allow phenotype prediction from molecular marker information and standard physiologically relevant environmental information.

Both LMMs and CGMs require extensive characterization of genotypes and environments. High-throughput technologies for genotyping and phenotyping provide new opportunities for upscaling phenotype prediction and increasing the response to selection in the breeding process.

3.1 Introduction

The target production area for most arable crops spans a range of environmental conditions. In the absence of diseases and pests (not considered in this review), local environmental conditions are a function of meteorological and soil variables on the one hand and management interventions on the other hand. These conditions will influence the phenotypic response of individual genotypes, and to some extent genotypes will create their 'own' environment, e.g. depending on how they use soil water across the season. The functional form by which environmental inputs are translated into phenotypes is sometimes referred to as the reaction norm (Woltereck 1909; Dobzhansky and Spassky 1963; Sarkar 1999; DeWitt and Scheiner 2004). Reaction norms depend both on environmental inputs and genetic factors. For a given (multi-locus) genotype, the reaction norm is the functional relationship between the phenotype and an environmental gradient, and is often linearised in some way. Modelling of the reaction norms for a set of genotypes is a central objective in many breeding and genetic studies. The prediction of the phenotypic response as a function of genetic and environmental factors is the basis for decisions that involve selection of superior genotypes for a defined environmental range (Hammer et al.2006; Chenu et al. 2011; Sadras et al. 2013).

Several important concepts in breeding and genetics have been defined in relation to the behaviour of reaction norms for a population of genotypes. Firstly, when

the reaction norms are non-constant, genotypes are said to show 'plasticity' (Bradshaw et al. 1965; DeWitt and Scheiner 2004; Sadras and Lawson 2011). Secondly, when the reaction norms for different genotypes are not parallel, this indicates the existence of genotype by environment interaction (GEI) (Finlay and Wilkinson 1963; van Eeuwijk et al. 2005). An extreme form of GEI is cross-over interaction, where the ranking of the genotypes varies with the environmental conditions (Baker 1988; Muir et al. 1992; Crossa et al. 2004). Another important concept in the context of the comparison of reaction norms is adaptation (Wright 1931, 1932; Finlay and Wilkinson 1963; Romagosa and Fox 1993; Cooper and Hammer 1996; Cooper 1999; Romagosa et al. 2013), i.e., some genotypes do better than other ones in a defined set of environmental conditions, the reaction norms of the adapted genotypes are then always above those of the less adapted. Finally, for a given genotype, 'stability' measures quantify the variation around the reaction norm (Lin and Binns 1988; Piepho 1998). So, while plasticity, GEI and adaptation refer to the expected response curve, which may be most simply thought of as the expectation in a linear regression model, stability refers to the variation around this expected response at a defined set of environmental conditions (Slafer and Kernich 1996; DeWitt and Scheiner 2004; van Eeuwijk et al. 2005; van Eeuwijk et al. 2010).

To select genotypes with superior average performance or a given degree of adaptation, predictions need to be made for the phenotype as a function of genotype and environment. These types of predictions occur at various stages in a breeding programme. In the early stages of breeding programmes, seed is limiting and large numbers of new genotypes produced as offspring from crosses between well-chosen parents are evaluated in one or a few trials, normally in small plots. For the earliest stages of a breeding programme, modelling of reaction norms is not possible and selection takes place on the mean performance. At intermediate stages, offspring populations are tested in a limited number of trials at various locations for one or a few years. In those cases when seed is still limiting, it is attractive to use partially replicated designs (Cullis et al. 2006; Smith et al. 2006) so that genotypes can be tested at a larger sample of environmental conditions. Selection can be done on the mean across trials, but there are also possibilities to select for adaptation. At the later stages, when there is sufficient seed for individual genotypes, a limited number of genotypes can be tested in a large number of trials, with again possibilities for selection on wide adaptation to a wide set of environments or narrow adaptation to a limited set of environments (Cooper et al. 2014). Simultaneously, at this stage selection on stability is possible.

When a population of genotypes is evaluated in multiple trials, reaction norms can be fitted to help in describing the observed data efficiently and to allow some form of selection on properties of the reaction norm. To evaluate the predictive quality of reaction norm models, special cross validation (CV) schemes have been proposed. In CV schemes, the data are subdivided in a training set, used to estimate model parameters, and a test set, used to assess prediction accuracy, which is the correlation between predicted and observed values (Meuwissen et al. 2001). For multiple environment data, various CV strategies have been proposed (Crossa et al. 2010, 2014; Burgueño et al. 2012; Heslot et al. 2012, 2013; Zhao et al. 2012). For a

transparent description of CV strategies, it is useful to introduce some notation. When genotypes were tested, evaluated or observed in at least one environment, we indicate this by the letter G. When this was not the case we use nG. For environments the same rule can be defined: E for observed environments, with at least one observed genotype, and nE for environments without observations (new environments). Specific combinations of genotype and environment can have been observed, GE, or not, nGE. Following this terminology, the set [G, E, GE] would indicate a genotype that was observed and an environment that was observed, while also the specific combination of genotype and environment was observed. The combination [G, E, nGE] indicates a genotype and environment that have been observed, but the specific combination of genotype and environment was not observed. This latter situation is typical for unbalanced genotype by environment data.

Figure 3.1 shows four scenarios that are relevant to prediction of phenotypes from genotypes and environments as well as to the calculation of accuracies and CV strategies. Scheme 1 pertains to situations in which both genotypes and environments were observed. Specific combinations of genotypes and environments may

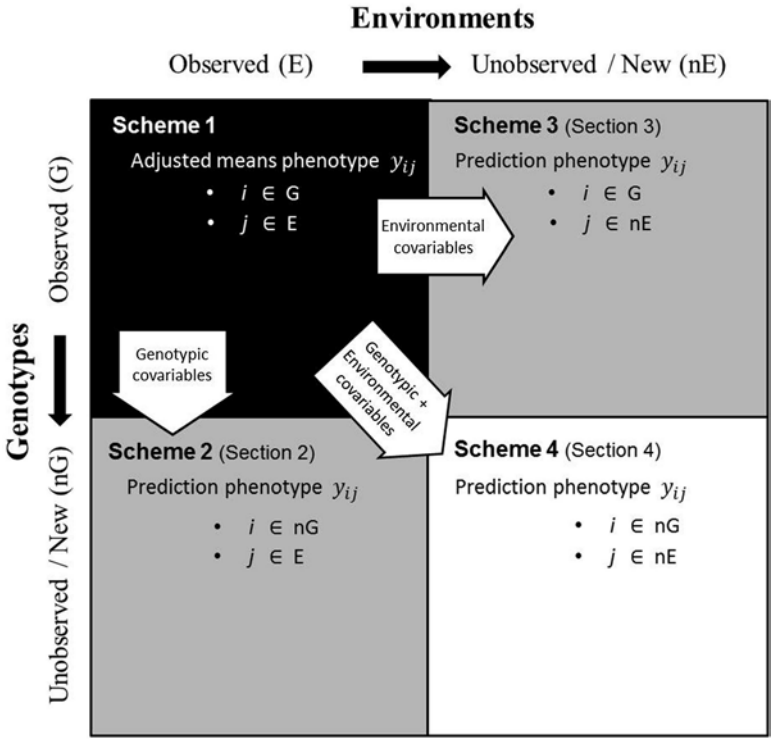


Fig. 3.1 Prediction scenarios, depending on whether genotypes were observed (G) or not observed (nG), and on whether environments were observed (E) or not observed (nE)

be present, [G, E, GE] or absent [G, E, nGE]. Phenotype predictions for Scheme 1 can be made by simple additive models. The Schemes 2, 3 and 4 are more interesting and we will concentrate on those. Potential strategies for assessment of accuracy in genomic prediction are predictions for new genotypes in observed environments [nG, E, nGE] (Scheme 2, Fig. 3.1); predictions for observed genotypes in new environments [G, nE, nGE] (Scheme 3, Fig. 3.1); and predictions for new genotypes in new environments [nG, nE, nGE] (Scheme 4, Fig. 3.1) (Utz et al. 2000; Calus and Veerkamp 2011; Burgueño et al. 2012; Schulz-Streeck et al. 2012; Guo et al. 2013; Crossa et al. 2014). Scheme 4 of CV obviously represents the strictest type of accuracy assessment. (For the notation, whenever nG or nE appears, necessarily nGE needs to appear as well, so for Schemes 2, 3 and 4, we can omit the specification nGE.)

To produce phenotype predictions for new genotypes (nG) from observed genotypes (G), it is essential to use statistical models that allow us to connect the new genotypes to the observed genotypes. The connections between nG and G can be achieved by the inclusion of explicit genotypic covariables in the statistical model, and/or by borrowing information via the correlation structure among genotypes, defined by their genetic similarities. Analogously, for predicting new environments, there needs to be a connection between nE and E via explicit environmental covariables and/or the correlation structure among environments. The latter correlation structure is an expression of environmental similarity as estimated from environmental characterizations.

In this chapter, we introduce linear mixed models (LMMs) as our default class of statistical prediction models. LMMs can be described as consisting of two parts: (1) a fixed part, corresponding to the mean; and (2) a random part defined by variances and covariances. Predictions in LMMs can be obtained via the fixed and the random part, although the statistical mechanism for prediction in those two cases is different. As an illustration, we provide an LMM for the phenotype of genotype i in environment j: $\underline{y}_{ij} = \mu_j + x_i\alpha_j + \beta_i z_j + \underline{GE}_{ij} + \underline{e}_{ij}$ (van Eeuwijk et al. 2010). The fixed part of this model is given by the expectation, or mean, for genotype i in environment j: $\mu_{ij} = \mu_j + x_i\alpha_j + \beta_i z_j$. Here μ_j is a fixed intercept (mean) for environment j, x_i is a genotypic covariable, for example a molecular marker, α_j is an environment specific slope corresponding to x_i. When x_i is a molecular marker, α_j is an environment specific quantitative trait locus (QTL) effect (Malosetti et al. 2004; Boer et al. 2007). For the environments, z_j is an environmental covariable, for example, a drought stress index, and β_i is a corresponding genotype specific slope, for example a genotype-specific sensitivity to drought stress.

For prediction via the fixed part, we use genotypic and/or environmental covariables as in classical regression (van Eeuwijk et al. 1996). Besides values for the covariable, x_i and z_j, prediction requires that we have estimates for the slopes, α_j and β_i. These can be obtained by fitting a model for the mean to training data, where we need to select suitable genotypic and/or environmental covariables. For prediction, we combine the estimated slopes in the training set with the values for genotypic and/or environmental covariables in the test set.

The random part of the model is determined by the terms $\underline{GE}_{ij}$ and $\underline{e}_{ij}$, the first term representing the (residual) genotypic effect of genotype *i* in environment *j*, the second term containing experimental (block) and measurement errors. (Random terms in model formulations are underlined.) The random terms are assumed to have a Gaussian distribution, with expectation zero and proper variance-covariance structures. The important random term for prediction purposes is $\underline{GE}_{ij}$. For this term, the correlations among genotypes on the one hand and the correlations among environments on the other hand determine the predictive properties of the LMM. Thus, for predictions via the random part of the LMM, correlations need to be estimated between observed and new genotypes (Scheme 2), observed and new environments (Scheme 3), or both (Scheme 4). Correlations among genotypes can be estimated from genotypic covariables, including molecular markers, and pedigree data, or a combination of genotypic covariables and pedigree. Correlations among environments follow from environmental covariables. Although important, we will largely ignore the error term $\underline{e}_{ij}$ in the remainder of this chapter. See Smith et al. (2001a) and Smith et al. (2005) for discussion on models for $\underline{e}_{ij}$.

The realization of the predictive potential of LMMs depends on the selection of genotypic covariables and environmental covariables, for the fixed part as well as for the random part. Physiological knowledge on genotypes and environments can help in the choice of covariables for inclusion in LMMs. For example, knowledge on the structure and use of crop growth models (CGMs) can help in the dissection of complex traits (Chapman et al. 2002b; Edmeades et al. 2004; Reynolds et al. 2009a), thereby suggesting genotypic and environmental covariables for inclusion in predictive LMMs. A CGM can suggest writing a complex target trait as a function of a set of simpler component traits and a set of environmental input variables (Yin et al. 2003, 2004; Chenu et al. 2008; Hammer et al. 2010). These component traits are traditionally related to physiological parameters in CGMs (see Chaps. 4, 5, 6, 7, 8, and 9 of this book). The CGM produces GEI as an emerging property of the interaction between the physiological parameters and the environmental information (Chapman et al. 2002a, 2008; Hammer et al. 2002, 2006, 2010). Interpreting the CGM as a function that transforms physiological parameters and environmental inputs into a complex trait, we can understand that when the CGM can be approximated by a linear function, the component traits may be entered as genotypic covariables and the environmental inputs as environmental covariables in an LMM for the complex trait.

In Sect. 3.2, we will discuss how statistical LMM models can be used to predict phenotypic responses for new genotypes in observed environments (Scheme 2; [nG, E, nGE]), observed genotypes in unobserved (new) environments (Scheme 3; [G, nE, nGE]), or new genotypes in new environments (Scheme 4; [nG, nE, nGE]). In Sect. 3.3, we will discuss the use of CGMs to predict the performance of genotypes for environments in which they were not tested. Section 3.4 will discuss the contribution of high throughput genotyping and phenotyping to models for phenotype prediction. Strategies to group genotypes and environments will also be discussed in this Section. We finish with some concluding remarks in Sect. 3.5.

3.2 Statistical Models to Predict Phenotypic Performance

Section 3.2.1 presents statistical models for predicting the phenotype of genotypes that were so far not tested in the environments for which we want to predict, although we do have information about these environments from phenotypic evaluations for other genotypes [nG, E, nGE], Scheme 2 in Fig. 3.1. The connection between observed genotypes (G) and not observed genotypes (nG) will come from explicit genotypic covariables and/or the genetic correlations among genotypes. Section 3.2.2 describes statistical models for predicting phenotypes in environments that were not used to test genotypes, although we do have phenotypic information about these genotypes in other environments [G, nE, nGE], Scheme 3 in Fig. 3.1. The connection between observed environments (E) and unobserved environment (nE), will result from the inclusion of explicit environmental covariables and/or the correlations among environments calculated on the basis of environmental characterizations. Section 3.2.3 discusses the most challenging prediction scenarios; predicting the phenotype of genotypes that were not tested so far, for environments that neither were tested [nG, nE, nGE], Scheme 4 in Fig. 3.1. Here, both explicit genotypic and environmental covariables are required for prediction.

3.2.1 Statistical Models to Predict Performance of Unobserved Genotypes in Observed Environments [nG, E, nGE]

Quantitative traits are determined by many loci, with allelic effects varying in magnitude. Specific genomic regions significantly associated with phenotypic variation may be identified as quantitative trait loci, or QTLs (see Chap. 1 of this book by Baldazzi et al.). Besides QTLs, or instead thereof, many other loci with small additive effects (polygenic effects) can contribute to phenotypic variation. None of these loci with small effects might by itself have an important phenotypic effect, but these loci together can still make a sizeable contribution to the phenotype. Model 3.1, includes loci with relatively large quantitative effects (QTLs) together with loci that have small effects.

$$\underline{y}_{ij}^{t} = \mu_j + \sum_{q=1}^{Q} x_{iq}\alpha_{jq} + \underline{G}_{ij} + \underline{e}_{ij} \tag{3.1}$$

In the multi-environment Model 3.1, y_{ij}^{t} represents the target trait, t, (for example, yield) of genotype i in environment j, μ_j is a fixed intercept term for each environment, x_{iq} is a genotypic covariable that represents DNA information of genotype i at QTL position q, and α_{jq} is the additive effect of the fixed QTL q in environment j. $\underline{G}_{ij}$ represents the residual genetic effect (polygenic effects) for genotype i in environment j. The matrix with elements $\underline{G}_{ij}$, $\left\{\underline{G}_{ij}\right\}$, has a multivariate normal distribution with zero mean, 0, and, as we will see later, a highly structured variance-covariance

matrix $\boldsymbol{\Sigma}$; $\left\{\underline{G}_{ij}\right\} \sim MVN\left(0, \Sigma\right)$. (For notational simplicity, we will omit the dimensions of the various matrices.) $\boldsymbol{\Sigma}$ defines the genetic variances and covariance for any two pairs of observations, $\underline{y}_{ij}^{t}$ and $\underline{y}_{i'j'}^{t}$, and depends on the genetic and environmental similarities of the two genotypes, *i* and *i'*, and the two environments *j* and *j'*. The term $\underline{e}_{ij}$ stands for a non-genetic residual, $\left\{\underline{e}_{ij}\right\} \sim MVN\left(0, \boldsymbol{R}\right)$, with $\boldsymbol{R}$ often allowing for specific residual variances per environment.

A simplification of Model 3.1 omits the genetic residual, $\underline{G}_{ij}$, and is appropriate when QTLs account for all of the genetic variation:

$$\underline{y}_{ij}^{t} = \mu + \sum_{q=1}^{Q} x_{iq}\alpha_{jq} + \underline{e}_{ij} \tag{3.2}$$

When Model 3.2 fits the data well, the performance of the unobserved genotype *i* in environment *j* can be predicted as;

$$\hat{\underline{y}}_{ij}^{t} = \hat{\underline{\mu}}_{j} + \sum_{q=1}^{Q} x_{iq}\hat{\underline{\alpha}}_{jq}$$

Compared with single-environment QTL models, multi-environment QTL models like Model 3.1 or Model 3.2 are more powerful in picking up QTLs and generally explain a larger amount of the genetic variance (Piepho 2000; Piepho and Möhring 2005; Mathews et al. 2008; Alimi et al. 2013). It has been shown that jointly considering multivariate phenotypes (i.e., the phenotype in multiple environments) allows a substantially greater separation between genotype classes than when considering univariate phenotypes (i.e., phenotype in a single environment) (Stephens 2013).

Another simplification of Model 3.1 occurs when we assume that there are no large discrete genetic effects in the form of QTLs that drive phenotypic differences, but that genetic effects are exclusively of a polygenic nature. A prediction model that generalizes the single environment genomic best linear unbiased prediction (G-BLUP) approach of (Meuwissen et al. 2001) to multi-environment prediction can be defined as:

$$\underline{y}_{ij}^{t} = \mu_{j} + \underline{G}_{ij} + \underline{e}_{ij} \tag{3.3}$$

In Model 3.3, the distribution of the polygenic effects $\underline{G}_{ij}$ is $\left\{\underline{G}_{ij}\right\} \sim MVN\left(0, \Sigma\right)$. Since $\boldsymbol{\Sigma}$ is a function of the genetic and environment similarities, the larger the similarity of unobserved genotypes with observed genotypes, and the larger the similarity of observed environments with unobserved environments, the more information is available for phenotype prediction, and the higher is the prediction accuracy (Crossa et al. 2006; Albrecht et al. 2014). Analogous to the classical partitioning of genetic and environmental effects, the covariance matrix $\boldsymbol{\Sigma}$ can be partitioned into a 'genotypic' variance-covariance matrix ($\boldsymbol{\Sigma}^{G}$), and an 'environmental' variance-covariance matrix ($\boldsymbol{\Sigma}^{E}$), such that $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{\boldsymbol{G}} \otimes \boldsymbol{\Sigma}^{E}$, i.e., the Kronecker product of the

genotypic variance-covariance matrix and the environmental variance-covariance matrix (West et al. 2006; Smith et al. 2005). It is important to realize that although $\boldsymbol{\Sigma}^E$ is called an 'environmental' variance-covariance matrix, $\boldsymbol{\Sigma}^E$ reflects genetic correlations among environments, and so plays a role in forming predictions in the multi-environment context. Examples of commonly used models for these two covariance matrices are given below.

$\boldsymbol{\Sigma}^G$ can be modelled as $\boldsymbol{\Sigma}^G = \boldsymbol{A}$, where $\boldsymbol{A}$ corresponds to the expected additive relationship matrix calculated from the coefficients of coancestry estimated from the pedigree, or to the realized additive relationship matrix estimated from molecular markers (Piepho et al. 2008). If the one step prediction with statistical models uses pedigree information, $\underline{G}_{ij}$ is commonly called "breeding value" (Falconer and Mackay 1996; Piepho et al. 2008). On the other hand, if the prediction uses molecular marker information, it is called "genomic estimated breeding value" (Burgueño et al. 2012; Piepho 2009).

In the multi-environment context, genotypic variances tend to change across environments with consequent changes in genotypic correlations for pairs of these environments. A flexible variance-covariance structure across environments $\boldsymbol{\Sigma}^E$, is required to achieve higher prediction accuracies. One flexible and parsimonious model for variances and covariances/correlations across environments is the factor analytic model (FA) (Table 3.1) (Smith et al. 2001a, 2005; Mathews et al. 2008).

The decision about when it is convenient to use Models 3.1, 3.2, or 3.3 depends on the genetic architecture of the target trait. If the trait is regulated by a few QTLs

Table 3.1 Variance-covariance models for the environmental covariance (Σ^E), ordered by increasing number of parameters. For simplicity, these examples assume three environments (m=3)

Name	Number of parameters	Structure
Identity	1	$\begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix}$
Compound symmetry	2	$\begin{bmatrix} \sigma^2+\varphi & \varphi & \varphi \\ \varphi & \sigma^2+\varphi & \varphi \\ \varphi & \varphi & \sigma^2+\varphi \end{bmatrix}$
Factor analytic, order 1	2 m	$\begin{bmatrix} \lambda_1^2+\psi_1 & \lambda_1\lambda_2 & \lambda_1\lambda_3 \\ \lambda_2\lambda_1 & \lambda_2^2+\psi_2 & \lambda_2\lambda_3 \\ \lambda_3\lambda_1 & \lambda_3\lambda_2 & \lambda_3^2+\psi_3 \end{bmatrix}$
Unstructured	m(m+1)/2	$\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{bmatrix}$

with large effects, a QTL model (Model 3.2) might provide the largest prediction accuracy. On the other hand, traits like grain yield, which are regulated by many genes with small effects might not show any significant QTL that can be included in Model 3.2. In this case, Model 3.3, whose predictions we will call GE-BLUPs because they can account for GEI, should integrate the large number of small additive effects into a multi-environment prediction model. For the intermediate case when traits have a few QTLs with large effects, and many other loci with very small additive effects, Model 3.1 is adequate. Bernardo (2014) suggested that it is convenient to consider QTLs (or genes) as fixed effects when they account for more than 10 % of the genetic variance. The simulations made by Bernardo (2014) show that the most adequate model depends on the genetic architecture of the trait, i.e., on the number of QTLs and the magnitudes of the QTL effects.

3.2.2 *Statistical Models to Predict Performance of Observed Genotypes in New Environments*

After genotypes have been phenotyped in some environments, it can be useful to predict their performance in other environments that were not used for evaluation. New environments could, for example, include future trials at known locations, which implies that none of the genotypes were observed in that environment yet [G, nE, nGE]. Thus, the correlation between observed environments and the predicted environments cannot be estimated from phenotypic data, or direct observations on the complex trait. In this case, we may use environmental covariables, like meteorological, soil or management covariables, as predictors in models for the mean or define correlations between environments in models for the variance-covariance structure.

Models for the mean that can be used to predict phenotypes in unobserved environments usually correspond to factorial regression models that incorporate environmental covariables. These models explicitly estimate the sensitivity of the QTL to environmental covariables (Model 3.4) (Campbell et al. 2004; Boer et al. 2007; Laperche et al. 2007; Malosetti et al. 2013; Romagosa et al. 2013). Hence, model parameters can have biological interpretation.

$$\underline{y}_{ij} = \mu_j + \sum_{q=1}^{Q} x_{iq}\left(\gamma_q + \delta_q z_j\right) + \underline{G}_{ij} + \underline{e}_{ij} \tag{3.4}$$

In Model 3.4, the additive effects (α_{jq}) of the fixed QTL q in environment j of Model 3.1 are replaced by a regression formulation, $\left(\gamma_q + \delta_q z_j\right)$, in which the effect of QTL q is a function of the environmental covariable z_j, and so changes over environments. When the covariable z_j is centered, the intercept term, γ_q, corresponds to the effect of the QTL in the average environment, while the slope δ_q corresponds to the sensitivity of the QTL q to the environmental covariable z_j. Although Model 3.4 does not explicitly restrict the environmental covariables to a particular range, it should

be considered that crops respond differently to covariables in the environmental extremes (e.g., too cold or too warm). So, the sensitivity of the genotype to the environmental covariables cannot be assumed constant outside the range of environments in which δ_q was estimated. A second issue that needs to be taken into account is that models like Model 3.4 do not make explicit in which phenological stage the environmental covariable is considered. Since the sensitivity of a crop to the environment varies throughout the development, environmental covariables included in the prediction model need to coincide with the developmental timing used to estimate the sensitivity.

For example, Boer et al. (2007) analysed grain yield and grain moisture for F5 maize testcross progenies evaluated across 12 environments in the U.S. corn belt. Since QTLs did not have a constant effect across environments (QTL by environment interaction), QTL effects were modelled conditional on longitude and year, both consequences of temperature differences during critical stages of the development. This factorial regression model allows prediction of yield and moisture at any location provided that temperatures during specific developmental stages are contained within those of the observed environments.

A second example is shown by Malosetti et al. (2004), who identified QTLs conferring differential sensitivity of grain yield to temperature during heading in a double haploid barley population. In a model like Model 3.4, the average daily temperature range during heading was the most important environmental covariable explaining differential QTL expression, i.e., the QTL allele from the parental line Steptoe conferred an extra grain yield of 0.112 t ha^{-1} for each extra degree Celsius during heading. Hence, yield could be predicted for unobserved environments if the average temperature for such environments was available. In that sense, Model 3.4 is closer to CGM than Model 3.1 because Model 3.4 explicitly represents environments on a continuous scale.

The second way to use environmental information for prediction is using environmental covariables to estimate similarities (covariances) among environments, analogous to the way molecular markers are used to characterize similarity among genotypes. If environmental covariables are considered as an indicator of environmental similarity, they can be used to estimate the environmental variance-covariance matrix in Model 3.3. Hence, $\boldsymbol{\Sigma}^E = \boldsymbol{\Omega}$, where $\boldsymbol{\Omega}$ is the variance-covariance matrix that accounts for the similarity in environmental conditions. The larger the covariance between observed and unobserved environments, the more information can be shared to make the predictions. The genotypic covariance $\boldsymbol{\Sigma}^G$ can be modelled as explained in Sect. 3.2.1 by imposing an additive relationship matrix to define $\boldsymbol{\Sigma}^G = \boldsymbol{A}$, where $\boldsymbol{A}$ can be estimated from the pedigree and/or from molecular markers.

Using multiple climatic variables to model the environmental covariance, as proposed by Jarquín et al. (2013) shows promise as a tool to predict genotypic performance in unobserved environments. However, many environmental covariables are correlated and not all need to be included in the model. Mechanistic CGMs such as APSIM have shown to be a good integrative tool to select subsets of variables that characterize environmental similarity (Chapman2008).

3.2.3 Statistical Models to Predict Performance of Unobserved Genotypes in New Environments

Section 3.2.1 presented models that used genotypic covariables to predict the phenotype on unobserved genotypes. Section 3.2.2 described how environmental covariables can be used in factorial regression models for prediction, and how to estimate the environmental covariance of a random term, necessary for prediction along the random part of an LMM. This Sect. 3.2.3 will combine both situations, aiming to predict the phenotype of genotypes that have not been tested yet for environments that have not been used for evaluation.

When predicting unobserved genotypes in new environments, both genotypic and environmental covariables are needed. In factorial regression-type of models, prediction of unobserved genotypes is possible, provided that the additive effects of each QTL allele can be estimated from the tested genotypes. The phenotypes of unobserved genotypes can also be predicted in new environments, provided that the sensitivity of the QTL effects along an environmental gradient (e.g., temperature), can be estimated from observed environments. In the example of Malosetti et al. (2004) presented in Sect. 3.2.2, phenotype prediction is possible for any environment provided the temperature remains within the range used to estimate the QTL sensitivity to temperature.

In models that entirely rely on the use of the variance-covariance structures imposed on genotypes and environments, prediction of unobserved genotypes in new environments is possible via the reconstruction of the full covariance matrix $\boldsymbol{\Sigma}$ from its components, $\boldsymbol{\Sigma}^G$ and $\boldsymbol{\Sigma}^E$. For the genotypic part, this runs via explicit pedigree information or information from genotypic covariables (molecular markers), while for the environmental part correlations between environments can be estimated from environmental characterization (temperature, precipitation, soil characteristics, etc.). Note that while in Sect. 3.2.1, $\boldsymbol{\Sigma}^G$ was calculated from genotypic covariables, and $\boldsymbol{\Sigma}^E$ was estimated from the phenotypic data on the target trait, here both $\boldsymbol{\Sigma}^G$ and $\boldsymbol{\Sigma}^E$ are estimated from explicit covariables.

3.3 Crop Growth Models to Predict Genotypic Performance

The algorithms in a CGM predict the target trait (e.g., grain yield) as a non-linear combination of underlying intermediate phenotypes (also commonly called “components”, e.g., biomass), which are calculated indirectly from a set of inputs to the CGM that typically comprise environment (soil, weather, and nutrients) data and CGM parameters derived from prior experimentation. GEI in the target trait is then a consequence of the interactions between the intermediate phenotypes (Chapman

et al. 2003; Tardieu 2003; Tardieu et al. 2005; Chenu et al. 2009; Makumburage et al. 2013).

Considering the CGM in reverse, we can state that the value of the target trait is able to be 'dissected' into these intermediate phenotypes (See Chap. 7 of this book by Hammer et al.). Although these intermediate phenotypes are likely to show less GEI than the target trait, they still correspond to an integration of genotypic responses to environmental conditions (e.g., they may show GEI). Ideally, a complete dissection of the target trait would comprise of a set of CGM input parameters that depend only on the genotype (for example, a genotypic sensitivity of development rate to the air temperature), and environmental covariables (Model 3.4), i.e., CGM parameters that do not show GEI (Slafer 2003; Yin et al. 2003; Bertin et al. 2010; Alam et al. 2014). The target trait for genotype *i* in environment *j* can be written as a function of CGM parameters and environmental inputs as follows:

$$\underline{y}_{ij}^{t} = \int f(\underline{\boldsymbol{y}}_{i}^{P}; \boldsymbol{z}_{j})\, dt + \underline{e}_{ij} \tag{3.5}$$

In Model 3.5, $\underline{y}_{ij}^{t}$ represents the target trait for genotype *i* in environment *j*, which is modelled as a function of multiple CGM parameters, $\underline{\boldsymbol{y}}_{i}^{P}$ (with *P* for parameter in the superscript), and multiple environmental inputs, $\boldsymbol{z}_{j}$, integrated over time (Fig. 3.2). The function *f* (;) embodies the algorithms that transform CGM parameters into intermediate phenotypes as well as the interactions between the intermediate phenotypes that lead to the target trait.

A commonly-studied CGM is APSIM, which currently has modules for several crops, e.g., wheat, canola, sorghum (Keating et al. 2003; Holzworth et al. 2014). In the case of APSIM-Wheat, growth (biomass accumulation) and development (phenological events, the functionality of plant structures or appearance of new structures) are calculated on a daily basis (Wang et al. 2002). The final phenotype (e.g., grain yield) is calculated as a function of a series of intermediate phenotypes. Examples of intermediate phenotypes are biomass, grain number and radiation interception on any given day or accumulated to a given day (Fig. 3.2). Intermediate phenotypes depend on CGM parameters that are genetically determined, and which modulate the phenotypic response to environmental covariables. Examples of CGM parameters are vernalization requirement and sensitivity to photoperiod, which are regulated by the *VRN* and the *PPD* alleles (Zheng et al. 2013).

CGM parameters, $\underline{\boldsymbol{y}}_{i}^{P}$, for phenotyped genotypes can be directly observed, estimated or calculated from the phenotypic measurements. However, given that CGM parameters depend on the genotype, they can also be predicted from genotypic covariables, i.e., molecular marker information. When we can identify the genetic basis of physiological parameters in terms of underlying QTLs, or, equivalently, when we can predict the physiological parameters from marker information, we can effectively predict the target trait from marker information and environmental inputs provided the intermediate traits and their interactions have been correctly identified

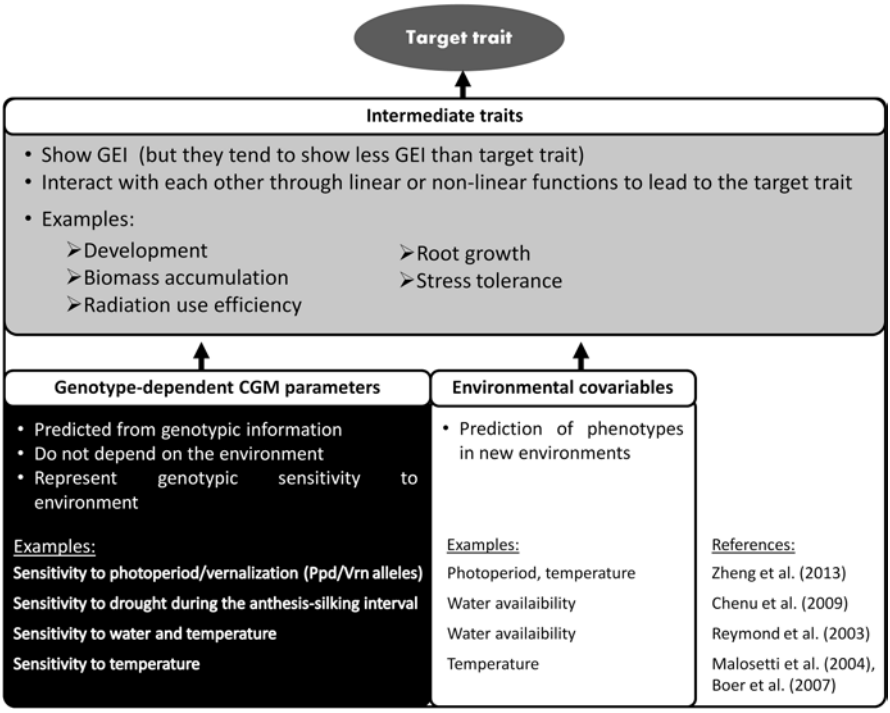


Fig. 3.2 Representation of the information flow in a CGM. The *black box* corresponds to CGM parameters that are dependent on the genotype, the *white box* represents environmental covariables and the *grey box* represents the intermediate phenotypes. Examples of different crops/traits/models are given for each category

and implemented in the CGM. Hence, predicted CGM parameters enable to predict the phenotype of genotypes that have not been observed yet. The prediction for individual CGM parameters ($\underline{y}_i^P$) would look like Model 3.6:

$$\underline{y}_i^P = \mu + \sum_{q=1}^{Q} x_{iq}\alpha_q + \underline{G}_i + \underline{e}_i \tag{3.6}$$

Like Model 3.1, Model 3.6 can be modified to include (i) only the QTLs, in a QTL model (Model 3.7) or (ii) only the polygenic effects ($\underline{G}_i$) , in a genomic prediction model with the random effects $\underline{G}_i$ being structured by a genetic relationship matrix (Model 3.8).

$$\underline{y}_i^P = \mu + \sum_{q=1}^{Q} x_{iq}\alpha_q + \underline{e}_i \tag{3.7}$$

$$\underline{y}_i^P = \mu + \underline{G}_i + \underline{e}_i \tag{3.8}$$

If more than one CGM parameter is to be predicted from molecular markers and/or pedigree information, Models 3.6, 3.7, and 3.8 could also be expanded to a multi-trait prediction model that takes into account possible correlations among the CGM parameters, in a model that is similar to the multi-environment Model 3.1. Modelling traits simultaneously allows to gain power for QTL detection and to detect QTLs with pleiotropic effects (Alimi et al. 2013; Stephens 2013).

Predictions for multiple CGM parameters, $\hat{\underline{y}}_i^P$, can be used as input in Model 3.5 to calculate intermediate phenotypes, and produce the prediction for the target trait, $\hat{\underline{\mu}}_{ij}^t$, in Model 3.9.

$$\hat{\underline{\mu}}_{ij}^t = \int f(\hat{\underline{y}}_i^P; \boldsymbol{z}_j)\, dt \tag{3.9}$$

In Model 3.9, the prediction accuracy of the target trait depends on the accuracy of the prediction of each of the components, and on the ability of the functions that transform CGM parameters into intermediate phenotypes to correctly describe the processes leading to the target trait.

CGMs with known/predicted genotypic parameters are a potentially useful tool to understand which traits can be advantageous in a given environment, and also to identify management practices that contribute to improved crop productivity (Yin et al. 2004; Hammer et al. 2006; Reynolds et al. 2009b; Harrison et al. 2014). In the context of adaptation to climate change, Zheng et al. (2012) modelled how phenology of current wheat varieties would influence their adaptation to future environments, which are expected to show different CO_2 and precipitation levels. In their second paper, Zheng et al. (2013) demonstrated that the flowering time of spring wheat genotypes can be modelled using the composition of their *VRN1* and *Ppd-D1* alleles together with responses derived from a single experiment with four environments: +/− treatments for vernalisation and extended photoperiod. Allelic combinations of loci *Vrn-A1*, *Vrn-B1*, *Vrn-D1*, and *Ppd-D1* were used to predict APSIM-wheat parameters of a population of genotypes. From a single experiment (replicated in 2 years), they validated the model with more than 250 wheat genotypes across the entire Australian wheat belt, and were able to simulate flowering time for any weather records in the wheat belt. These conclusions can be useful to guide breeders in the process of determining which alleles should be considered in the selection process.

Bogard et al. (2014) extended this approach further to model the drivers of flowering time in winter wheat as functions of major genes as well as SNPs derived from association mapping, i.e., allowing prediction of unknown genotypes (but with known genes and SNPs) in new environments. In both Zheng et al. (2013) and

Bogard et al. (2014), the predictions for heading date using the gene-based predicted parameters corresponded well with the observed dates to heading. Attributes that contributed to a successful phenotype prediction were (i) a well-defined CGM for heading date (Slafer and Rawson 1994), (ii) a well-defined set of environmental covariables with corresponding CGM parameters, and (iii) a genetic basis of the CGM parameters (Snape et al. 2001).

The same approach has also shown to be successful for other more complex and less heritable traits such as grain yield under drought. For example, Chenu et al. (2009) used APSIM to model the impact of QTLs controlling the intermediate traits leaf and silk elongation on maize grain yield. The intercept and slope of these intermediate traits in response to meristem temperature, evaporative demand and soil water deficit were genotype-dependent (Reymond et al. 2003, 2004).

Unfortunately, the identification of CGM parameters is sometimes less straightforward for complex traits like grain yield. Yin et al. (2000) showed an example in barley with a successful estimation of QTL effects for the CGM parameters, but with a poor prediction of grain yield. The correlation between the observed CGM parameters, i.e. phenotype of CGM parameters, and the QTL predictions of the same parameters was high. However, the correlation between yield predictions of the CGM, whether phenotype based or QTL-prediction based, and observed yield was not high. The cause of the poor predictions did not reside in the fact that the CGM parameters were replaced by predicted parameters from the QTL model, but in the fact that the CGM was unable to predict yield from its component traits. Similar work has been recently reported by Gu et al. (2014) on grain yield of rice crop, using a new CGM, which gave more promising results. However, efforts to improve CGM for predicting complex traits like grain yield are still strongly needed.

The example from Yin et al. (2000) shows that although the integrated statistical and CGM modelling allows for a larger flexibility, it might result in more complex and fragile models, because the approach can break down at the level of the estimation of the CGM parameters and at the level of the integration of these CGM parameters to calculate the intermediate phenotypes. However, even if the CGMs are not fully able to predict the target trait, it is valuable to develop models of intermediate traits as well as for yield per se. Breeders can still be interested to recombine lines with high levels of proven intermediate traits with the expectation that these should on average result in better yield when further crossing and selection is done, i.e., because the selection on intermediate traits should already have improved part of the physiological adaptation pathway (Cooper et al. 2014). If breeders select mainly on yield per se, then it may be less likely that selected genotypes will also have high radiation use efficiency (RUE) or transpiration efficiency (TE) or traits for which genetic variation was not expressed in the given selection environment.

The examples of Zheng et al. (2013), Bogard et al. (2014), and Chenu et al. (2009) show that CGMs are a tool to integrate complex information from the genotypic, organ, and crop level (see also examples reviewed in Chap. 9 of this book by Yin et al.). Dissection of a target trait into component traits at different levels of biological organisation allows phenotype prediction for the target trait in the face of genotype by environment and QTL by environment interactions for that same trait.

Hence, the combined approach of statistical QTL modelling and CGM is an alternative to model complex GEI interactions (Yin et al. 2004).

3.4 Further Considerations

3.4.1 Classification of Environments

Sections 3.2.2 and 3.2.3 presented models to predict the performance of genotypes in new environments ([G, nE, nGE] or [nG, nE, nGE]). However, if there are repeatable patterns that allow to classify environments, these patterns might help to reduce the complexity of $\boldsymbol{\Sigma}^E$ and thereby improve the accuracy of prediction.

One example of repeatable patterns that often justifies to group environments is the presence of regions. Here, we understand 'regions' (or mega-environments) as a group of locations where genotypes perform consistently across years (Bull et al. 1992; Gauch and Zobel 1997; Basford and Cooper1998; Yan et al. 2000). Environments inside the same region are expected to be more homogeneous in terms of genotypic ranking, i.e., less GEI inside the regions (e.g. Atlin et al. 2000; Burgueño et al. 2008). In dryland production areas, other groupings may relate to characteristics of the soil (shallow/deep, low/high water holding capacity) and the management of the crop (sowing date, row spacing arrangement, etc.). De la Vega and Chapman (2010) showed how multiple component traits related to yield for a complex set of mega-environments in Argentina.

If locations can be grouped into regions, it is generally convenient to breed for specific adaptation to those regions, instead of broad adaptation across regions (Atlin et al. 2000, 2011). In this case, predictions can be produced for the whole of a region, or for new environments within a region. Precision of yield estimates might still benefit from the information of neighbouring regions by means of the covariance structure in a mixed model (Piepho and Möhring 2005; Kleinknecht et al. 2013).

When phenotypes are not available for all the locations of interest, environmental covariables can also be used to classify environments, and reduce the complexity of $\boldsymbol{\Sigma}^E$. Classifying environments into regions on the basis of environmental similarity, potentially allows to (i) predict new environments (as discussed in Sects. 3.2.2 and 3.2.3), and also (ii) define the target population of environments, where a particular genotype is to be grown (Chapman et al. 2000a; Hammer et al. 2002; Chenu et al. 2011). CGMs are a powerful tool to identify relevant environmental factors (Chapman 2008; Messina et al. 2011), and the periods when the crop is most sensitive to those factors (Chenu et al. 2013). For example, considering drought seasonal patterns could give a better indication of the environment types, instead of the total rainfall per year (Chapman et al. 2000a, b).

A further application of explicit environmental characterization is to weight environments based on their expected relevance for future years (Podlich et al.

1999). This means that environmental conditions that are more likely to occur receive more weight when doing the predictions, compared to less likely environmental conditions.

3.4.2 Population Structure

Sections 3.2 and 3.3 discussed different models to predict phenotypes of unobserved genotypes using molecular marker information. In those sections, $\boldsymbol{\Sigma}^G$ had the structure of the genomic relationship matrix, without explicitly specifying sub-populations. However, genetic relatedness between training and test sets largely influences prediction accuracy (Windhausen et al. 2012; Riedelsheimer et al. 2013). Hence, when there is strong population structure, it is necessary to define whether prediction will be done among or within populations. When predictions are limited to specific sub-populations, accuracy is commonly larger than when predicting across sub-populations, or when correcting for population structure (Daetwyler et al. 2012; Guo et al. 2014).

Methods to consider population structure in the model for genomic prediction can be based on the incorporation of the eigenvectors of the genotype by molecular marker data matrix (Patterson et al. 2006; Janss et al. 2012). Another option is to consider population structure in the design of the cross-validation scheme, for example by a stratified cross-validation design conditional on known population structure to ensure that each sub-population is equally represented in the training and validation sets (Albrecht et al. 2014; Guo et al. 2014).

3.4.3 Next Generation Sequencing

With the recent development of next generation sequencing technologies, genotyping costs have been largely reduced, allowing improving the genotypic characterization of important crops as barley, wheat and potato (Poland et al. 2012b; Uitdewilligen et al. 2013). In sequence-based genotyping approaches, marker discovery and genotyping are completed at the same time, allowing for faster genotyping processes (Poland and Rife 2012). The shorter time needed is thanks to the combination of restriction enzymes, sequencing, imaging, and genome alignment and assembly methods (Metzker 2010; Elshire et al. 2011).

These technologies permit the genotyping of larger populations of plants with higher marker density and increased mapping resolution (Varshney et al. 2014). Larger marker density increases the chances of including causal loci that otherwise would not have been considered in models for phenotype prediction (Spindel et al. 2013). More loci in the model means increased genomic prediction accuracy (Poland et al. 2012a). However, models for phenotype prediction have diminishing

returns on additional markers once the point of "marker saturation" has been reached, which depends on the genetic diversity of the population (Jannink et al. 2010; Heffner et al. 2011; Poland et al. 2012a).

Other questions regarding larger numbers of markers that remain not fully answered are: (i) how imputation of missing genotype data or haplotype inferences may affect prediction accuracies when genotyping by sequencing is used (Crossa et al. 2013), (ii) how to reduce the computational time needed because of the large number of markers (Verbyla and Cullis 2012), and (iii) how to improve model diagnostics, distinguishing between loci with large effects, and loci with smaller effects (Bernardo 2014).

3.4.4 High-Throughput Phenotyping to Input to Models for Phenotype Prediction

Mixed models and CGM discussed in Sects. 3.2 and 3.3 are promising tools for phenotype prediction. However, these models require the phenotyping of multiple genotypes, traits and environments. With the reduction of genotyping costs, evaluating the populations phenotypically has become the limiting factor (Cobb et al. 2013).

High-throughput phenotyping platforms can either measure the target trait directly, or measure one or more traits that are correlated with the target trait. The use of CGMs allows estimation of hard-to-measure traits such as seasonal water use, given inputs of leaf area over time and canopy thermal characteristics, for example. Correlated traits measured by high-throughput phenotyping platforms can be used as inputs in models like Model 3.1. To do so, traits must: (i) have high genetic correlation with the target trait in the target environment, (ii) be less affected by environment (have a larger heritability) than the target trait, and (iii) provide an easy and reliable measurement, which is less expensive than the target trait itself (Bänziger 2000; Araus et al. 2008; Prasanna et al. 2013). When measuring correlated traits, high-throughput phenotyping platforms could be particularly useful for obtaining detailed non-destructive measurements of plant characteristics that collectively provide reliable estimates of trait phenotypes (Cabrera-Bosquet et al. 2012; Prasanna et al. 2013; Cooper et al. 2014).

High-throughput phenotyping platforms are commonly used under two scenarios: (i) precise phenotyping under controlled environments that aim at representing different levels of environmental quality, and (ii) phenotyping in environments that correspond to a sample of environments in the field. The main advantage of controlled environments is that screening protocols can be more easily standardized, ensuring that plants are exposed to fairly reliable levels of stress. Hence, controlled environments offer the stability to search for attractive phenotypes or genotypes in a specific type of environment, e.g. drought stress (Cobb et al. 2013; Passioura 2012).

Growth under controlled conditions is usually different from that under field conditions. Hence, high-throughput phenotyping platforms in controlled environments might not lead to the identification of important yield-determining processes and promising genotypes in the field (Passioura 2012). This limits the application for phenotyping to specific stages of the crop (e.g., early vigour), or to traits that are correlated with the target trait (e.g., carbon isotope discrimination as an indicator of water use efficiency (Passioura 2012; Prasanna et al. 2013).

Popular high-throughput phenotyping techniques are those based on spectral technologies or remote sensing, such as near infrared spectroscopy (NIRS), or image analysis. These techniques are a powerful tool that can provide information about multiple traits from only one or few images, and can be applied in controlled conditions as well as in field trials.

One example of how phenotypes obtained by image analysis can be included in phenotype prediction is shown by van der Heijden et al. (2012). Here, QTLs for leaf area were identified from the 3D representation of the plant canopy reconstructed from stereo images. The QTLs for leaf area from the image analysis agreed with the QTLs detected when using manually measured leaf areas, showing the potential of stereo images to characterize phenotypically breeding populations.

Image analysis introduces potentially larger measurement errors than conventional measurement techniques. For that reason, image information should be first carefully selected with the aid of statistical and physiological knowledge, in an automatized and standardized fashion, before incorporating it in the genetic analysis (Eberius and Lima-Guerra 2009; Hartmann et al. 2011). Hence, models accounting separately for the measurement error and for the experimental (plot) error should be considered (Smith et al. 2001b).

3.5 Concluding Remarks

This chapter discussed several approaches that aim at predicting the phenotype in a multi-environment context. These approaches ranged from pure statistical models and pure CGMs, to a combination of both types of models. Special attention was given to different prediction scenarios; unobserved genotypes, new environments, and the combination of both. How prediction accuracy can profit from the large availability of environmental and genotypic information was also discussed, aiming at integrating physiological and statistical knowledge. Phenotypic and genomic data start to become abundant. The challenge for better phenotype prediction and more effective selection lies in more sophisticated procedures for selection of genotypic and environmental covariables in models for phenotype prediction, separating the signal from the noise.

Acknowledgements Daniela Bustos-Korts thanks Becas Chile (CONICYT) for the financial support in form of a PhD scholarship. Marcos Malosetti and Fred van Eeuwijk worked on this chapter as part of a project financed by the Generation Challenge Program – Integrated Breeding Platform (https://www.integratedbreeding.net/).

References

Alam MM, Mace ES, van Oosterom EJ, Cruickshank A, Hunt CH, Hammer GL, Jordan DR (2014) QTL analysis in multiple sorghum populations facilitates the dissection of the genetic and physiological control of tillering. TAG Theor App Genet (Theor angew Genet) 127(10):2253–2266. doi:10.1007/s00122-014-2377-9

Albrecht T, Auinger H-J, Wimmer V, Ogutu J, Knaak C, Ouzunova M, Piepho H-P, Schön C-C (2014) Genome-based prediction of maize hybrid performance across genetic groups, testers, locations, and years. TAG Theor App Genet (Theor angew Genet) 127:1375–1386. doi:10.1007/s00122-014-2305-z

Alimi NA, Bink MCAM, Dieleman JA, Magán JJ, Wubs AM, Palloix A, van Eeuwijk FA (2013) Multi-trait and multi-environment QTL analyses of yield and a set of physiological traits in pepper. Theor Appl Genet 126(10):2597–2625. doi:10.1007/s00122-013-2160-3

Araus JL, Slafer GA, Royo C, Serret MD (2008) Breeding for yield potential and stress adaptation in cereals. Crit Rev Plant Sci 27(6):377–412. doi:10.1080/07352680802467736

Atlin GN, Baker RJ, McRae KB, Lu X (2000) Selection response in subdivided target regions. Crop Sci 40(1):7–13. doi:10.2135/cropsci2000.4017

Atlin GN, Kleinknecht K, Singh GP, Piepho HP (2011) Managing genotype x environment interaction in plant breeding programs: a selection theory approach. J Indian Soc Agric Stat 65(2):237–247

Baker RJ (1988) Tests for crossover genotype-environmental interactions. Can J Plant Sci 68(2):405–410. doi:10.4141/cjps88-051

Bänziger M (2000) Breeding for drought and nitrogen stress tolerance in maize: from theory to practice. Cimmyt, Mexico

Basford KE, Cooper M (1998) Genotype x environment interactions and some considerations of their implications for wheat breeding in Australia. Aust J Agric Res 49(2):153–174. doi:10.1071/A97035

Bernardo R (2014) Genomewide selection when major genes are known. Crop Sci 54(1):68–75. doi:10.2135/cropsci2013.05.0315

Bertin N, Martre P, Génard M, Quilot B, Salon C (2010) Under what circumstances can process-based simulation models link genotype to phenotype for complex traits? Case-study of fruit and grain quality traits. J Exp Bot 61(4):955–967. doi:10.1093/jxb/erp377

Boer MP, Wright D, Feng L, Podlich DW, Luo L, Cooper M, van Eeuwijk FA (2007) A mixed-model quantitative trait loci (QTL) analysis for multiple-environment trial data using environmental covariables for QTL-by-environment interactions, with an example in maize. Genetics 177(3):1801–1813. doi:10.1534/genetics.107.071068

Bogard M, Ravel C, Paux E, Bordes J, Balfourier F, Chapman SC, Le Gouis J, Allard V (2014) Predictions of heading date in bread wheat (Triticum aestivum L.) using QTL-based parameters of an ecophysiological model. J Exp Bot 65(20):5849–5865. doi:10.1093/jxb/eru328

Bradshaw AD, Caspari EW, Thoday JM (1965) Evolutionary significance of phenotypic plasticity in plants. Adv Genet 13:115–155

Bull JK, Cooper M, DeLacy IH, Basford KE, Woodruff DR (1992) Utility of repeated checks for hierarchical classification of data from plant breeding trials. Field Crops Res 30((1–2)):79–95. doi:10.1016/0378-4290(92)90058-H

Burgueño J, Crossa J, Cornelius PL, Yang R-C (2008) Using factor analytic models for joining environments and genotypes without crossover genotype×environment interaction. Crop Sci 48(4):1291–1305. doi:10.2135/cropsci2007.11.0632

Burgueño J, De los Campos G, Weigel K, Crossa J (2012) Genomic prediction of breeding values when modeling genotype×environment interaction using pedigree and dense molecular markers. Crop Sci 52(2):707–719. doi:10.2135/cropsci2011.06.0299

Cabrera-Bosquet L, Crossa J, von Zitzewitz J, Serret MD, Araus JL (2012) High-throughput phenotyping and genomic selection: the frontiers of crop breeding converge. J Integr Plant Biol 54(5):312–320. doi:10.1111/j.1744-7909.2012.01116.x

Calus M, Veerkamp R (2011) Accuracy of multi-trait genomic selection using different methods. Genet Sel Evol 43(1):26

Campbell BT, Baenziger PS, Eskridge KM, Budak H, Streck NA, Weiss A, Gill KS, Erayman M (2004) Using environmental covariates to explain genotype × environment and QTL × environment interactions for agronomic traits on chromosome 3A of wheat. Crop Sci 44(2):620–627. doi:10.2135/cropsci2004.6200

Chapman S (2008) Use of crop models to understand genotype by environment interactions for drought in real-world and simulated plant breeding trials. Euphytica 161(1–2):195–208. doi:10.1007/s10681-007-9623-z

Chapman SC, Cooper M, Hammer GL, Butler DG (2000a) Genotype by environment interactions affecting grain sorghum. II. Frequencies of different seasonal patterns of drought stress are related to location effects on hybrid yields. Aust J Agric Res 51(2):209–222. doi:10.1071/AR99021

Chapman SC, Hammer GL, Butler DG, Cooper M (2000b) Genotype by environment interactions affecting grain sorghum. III. Temporal sequences and spatial patterns in the target population of environments. Aust J Agric Res 51(2):223–234. doi:10.1071/AR99022

Chapman SC, Cooper M, Hammer GL (2002a) Using crop simulation to generate genotype by environment interaction effects for sorghum in water-limited environments. Aust J Agric Res 53(4):379–389. doi:10.1071/AR01070

Chapman SC, Hammer GL, Podlich DW, Cooper M (2002b) Linking bio-physical and genetic models to integrate physiology, molecular biology and plant breeding. In: Kang M (ed) Quantitative genetics, genomics, and plant breeding. CAB International, Wallingford, pp 167–187

Chapman S, Cooper M, Podlich D, Hammer G (2003) Evaluating plant breeding strategies by simulating gene action and dryland environment effects. Agron J 95(1):99–113. doi:10.2134/agronj2003.9900

Chenu K, Chapman SC, Hammer GL, McLean G, Salah HBH, Tardieu F (2008) Short-term responses of leaf growth rate to water deficit scale up to whole-plant and crop levels: an integrated modelling approach in maize. Plant Cell Environ 31(3):378–391. doi:10.1111/j.1365-3040.2007.01772.x

Chenu K, Chapman SC, Tardieu F, McLean G, Welcker C, Hammer GL (2009) Simulating the yield impacts of organ-level quantitative trait loci associated with drought response in maize: a "gene-to-phenotype" modeling approach. Genetics 183(4):1507–1523. doi:10.1534/genetics.109.105429

Chenu K, Cooper M, Hammer GL, Mathews KL, Dreccer MF, Chapman SC (2011) Environment characterization as an aid to wheat improvement: interpreting genotype–environment interactions by modelling water-deficit patterns in North-Eastern Australia. J Exp Bot 62(6):1743–1755. doi:10.1093/jxb/erq459

Chenu K, Deihimfard R, Chapman SC (2013) Large-scale characterization of drought pattern: a continent-wide modelling approach applied to the Australian wheatbelt – spatial and temporal trends. New Phytol 198(3):801–820. doi:10.1111/nph.12192

Cobb JN, Declerck G, Greenberg A, Clark R, McCouch S (2013) Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. Theor Appl Genet 126(4):867–887. doi:10.1007/s00122-013-2066-0

Cooper M (1999) Concepts and strategies for plant adaptation research in rainfed lowland rice. Field Crops Res 64(1–2):13–34. doi:10.1016/S0378-4290(99)00048-9

Cooper M, Hammer GL (1996) Plant adaptation and crop improvement. CAB International, Wallingford

Cooper M, Messina CD, Podlich D, Totir LR, Baumgarten A, Hausmann NJ, Wright D, Graham G (2014) Predicting the future of plant breeding: complementing empirical evaluation with genetic prediction. Crop Pasture Sci 65(4):311–336

Crossa J, Yang R-C, Cornelius P (2004) Studying crossover genotype × environment interaction using linear-bilinear models and mixed models. JABES 9(3):362–380. doi:10.1198/108571104x4423

Crossa J, Burgueño J, Cornelius PL, McLaren G, Trethowan R, Krishnamachari A (2006) Modeling genotype × environment interaction using additive genetic covariances of relatives for predicting breeding values of wheat genotypes. Crop Sci 46(4):1722–1733. doi:10.2135/cropsci2005.11-0427

Crossa J, de los Campos G, Perez P, Gianola D, Burgueño J, Araus JL, Makumbi D, Singh RP, Dreisigacker S, Yan JB, Arief V, Banziger M, Braun HJ (2010) Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics 186(2):713-U406. doi:10.1534/genetics.110.118521

Crossa J, Beyene Y, Kassa S, Pérez P, Hickey JM, Chen C, de los Campos G, Burgueño J, Windhausen VS, Buckler E, Jannink J-L, Lopez Cruz MA, Babu R (2013) Genomic prediction in maize breeding populations with genotyping-by-sequencing. G3 Genes Genomes Genet 3(11):1903–1926. doi:10.1534/g3.113.008227

Crossa J, Perez P, Hickey J, Burgueno J, Ornella L, Ceron-Rojas J, Zhang X, Dreisigacker S, Babu R, Li Y, Bonnett D, Mathews K (2014) Genomic prediction in CIMMYT maize and wheat breeding programs. Heredity 112(1):48–60. doi:10.1038/hdy.2013.16

Cullis BR, Smith AB, Coombes NE (2006) On the design of early generation variety trials with correlated data. J Agric Biol Environ Stat 11(4):381–393. doi:10.1198/108571106x154443

Daetwyler HD, Kemper KE, van der Werf JHJ, Hayes BJ (2012) Components of the accuracy of genomic prediction in a multi-breed sheep population. J Anim Sci 90(10):3375–3384. doi:10.2527/jas.2011-4557

De la Vega AJ, Chapman SC (2010) Mega-environment differences affecting genetic progress for yield and relative value of component traits. Crop Sci 50(2):574–583. doi:10.2135/cropsci2009.04.0209

DeWitt TJ, Scheiner SM (2004) Phenotypic plasticity: functional and conceptual approaches. Oxford University Press, Oxford

Dobzhansky T, Spassky B (1963) Genetics of natural populations. XXXIV. Adaptive norm, genetic load and genetic elite in drosophila pseudoobscura. Genetics 48(11):1467–1485

Eberius M, Lima-Guerra J (2009) High-throughput plant phenotyping – data acquisition, transformation, and analysis. In: Edwards D, Stajich J, Hansen D (eds) Bioinformatics. pp 259–278. doi:10.1007/978-0-387-92738-1_13

Edmeades GO, McMaster GS, White JW, Campos H (2004) Genomics and the physiologist: bridging the gap between genes and crop response. Funct Plant Biol 90(1):5–18. doi:10.1016/j.fcr.2004.07.002

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6(5), e19379. doi:10.1371/journal.pone.0019379

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics. 4th edn. Harlow, GB: Longman

Finlay K, Wilkinson G (1963) The analysis of adaptation in a plant-breeding programme. Aust J Agric Res 14(6):742–754. doi:10.1071/AR9630742

Gauch HG, Zobel RW (1997) Identifying mega-environments and targeting genotypes. Crop Sci 37(2):311–326. doi:10.2135/cropsci1997.0011183X003700020002x

Gu J, Yin X, Zhang C, Wang H, Struik PC (2014) Linking ecophysiological modelling with quantitative genetics to support marker-assisted crop design for improved yields of rice (Oryza sativa) under drought stress. Ann Bot 114(3):499–511. doi:10.1093/aob/mcu127

Guo Z, Tucker DM, Wang D, Basten CJ, Ersoz E, Briggs WH, Lu J, Li M, Gay G (2013) Accuracy of across-environment genome-wide prediction in maize nested association mapping populations. G3 Genes Genomes Genet 3(2):263–272. doi:10.1534/g3.112.005066

Guo Z, Tucker D, Basten C, Gandhi H, Ersoz E, Guo B, Xu Z, Wang D, Gay G (2014) The impact of population structure on genomic prediction in stratified populations. Theor Appl Genet 127(3):749–762. doi:10.1007/s00122-013-2255-x

Hammer GL, Kropff MJ, Sinclair TR, Porter JR (2002) Future contributions of crop modelling – from heuristics and supporting decision making to understanding genetic regulation and aiding crop improvement. Eur J Agron 18(1–2):15–31. doi:10.1016/S1161-0301(02)00093-X

Hammer G, Cooper M, Tardieu F, Welch S, Walsh B, van Eeuwijk F, Chapman S, Podlich D (2006) Models for navigating biological complexity in breeding improved crop plants. Trends Plant Sci 11(12):587–593. doi:10.1016/j.tplants.2006.10.006

Hammer GL, van Oosterom E, McLean G, Chapman SC, Broad I, Harland P, Muchow RC (2010) Adapting APSIM to model the physiology and genetics of complex adaptive traits in field crops. J Exp Bot 61(8):2185–2202. doi:10.1093/jxb/erq095

Harrison MT, Tardieu F, Dong Z, Messina CD, Hammer GL (2014) Characterizing drought stress and trait influence on maize yield under current and future conditions. Glob Chang Biol 20(3):867–878. doi:10.1111/gcb.12381

Hartmann A, Czauderna T, Hoffmann R, Stein N, Schreiber F (2011) HTPheno: an image analysis pipeline for high-throughput plant phenotyping. BMC Bioinf 12(1):148

Heffner EL, Jannink J-L, Sorrells ME (2011) Genomic selection accuracy using multifamily prediction models in a wheat breeding program. Plant Genome 4(1):65–75. doi:10.3835/plantgenome2010.12.0029

Heslot N, Yang H-P, Sorrells ME, Jannink J-L (2012) Genomic selection in plant breeding: a comparison of models. Crop Sci 52(1):146–160. doi:10.2135/cropsci2011.06.0297

Heslot N, Jannink J-L, Sorrells ME (2013) Using genomic prediction to characterize environments and optimize prediction accuracy in applied breeding data. Crop Sci 53(3):921–933. doi:10.2135/cropsci2012.07.0420

Holzworth DP, Huth NI, deVoil PG, Zurcher EJ, Herrmann NI, McLean G, Chenu K, van Oosterom EJ, Snow V, Murphy C, Moore AD, Brown H, Whish JPM, Verrall S, Fainges J, Bell LW, Peake AS, Poulton PL, Hochman Z, Thorburn PJ, Gaydon DS, Dalgliesh NP, Rodriguez D, Cox H, Chapman S, Doherty A, Teixeira E, Sharp J, Cichota R, Vogeler I, Li FY, Wang E, Hammer GL, Robertson MJ, Dimes JP, Whitbread AM, Hunt J, van Rees H, McClelland T, Carberry PS, Hargreaves JNG, MacLeod N, McDonald C, Harsdorf J, Wedgwood S, Keating BA (2014) APSIM – evolution towards a new generation of agricultural systems simulation. Environ Model Software. doi:10.1016/j.envsoft.2014.07.009

Jannink J-L, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. Brief Funct Genomics 9(2):166–177. doi:10.1093/bfgp/elq001

Janss L, De los Campos G, Sheehan N, Sorensen D (2012) Inferences from genomic models in stratified populations. Genetics 192(2):693–704. doi:10.1534/genetics.112.141143

Jarquín D, Crossa J, Lacaze X, Cheyron P, Daucourt J, Lorgeou J, Piraux F, Guerreiro L, Pérez P, Calus M, Burgueño J, Campos G (2013) A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theor Appl Genet 3:1–13. doi:10.1007/s00122-013-2243-1

Keating BA, Carberry PS, Hammer GL, Probert ME, Robertson MJ, Holzworth D, Huth NI, Hargreaves JNG, Meinke H, Hochman Z, McLean G, Verburg K, Snow V, Dimes JP, Silburn M, Wang E, Brown S, Bristow KL, Asseng S, Chapman S, McCown RL, Freebairn DM, Smith CJ (2003) An overview of APSIM, a model designed for farming systems simulation. Eur J Agron 18(3–4):267–288. doi:10.1016/S1161-0301(02)00108-9

Kleinknecht K, Möhring J, Singh KP, Zaidi PH, Atlin GN, Piepho HP (2013) Comparison of the performance of best linear unbiased estimation and best linear unbiased prediction of genotype effects from zoned Indian maize data. Crop Sci 53(4):1384–1391. doi:10.2135/cropsci2013.02.0073

Laperche A, Brancourt-Hulmel M, Heumez E, Gardet O, Hanocq E, Devienne-Barret F, Le Gouis J (2007) Using genotype × nitrogen interaction variables to evaluate the QTL involved in wheat tolerance to nitrogen constraints. Theor Appl Genet 115(3):399–415. doi:10.1007/s00122-007-0575-4

Lin CS, Binns MR (1988) A method of analyzing cultivar x location x year experiments: a new stability parameter. Theor Appl Genet 76(3):425–430. doi:10.1007/bf00265344

Makumburage GB, Richbourg HL, LaTorre KD, Capps A, Chen C, Stapleton AE (2013) Genotype to phenotype maps: multiple input abiotic signals combine to produce growth effects via attenuating signaling interactions in maize. G3 Genes Genomes Genet 3(12):2195–2204. doi:10.1534/g3.113.008573

Malosetti M, Voltas J, Romagosa I, Ullrich SE, van Eeuwijk FA (2004) Mixed models including environmental covariables for studying QTL by environment interaction. Euphytica 137(1):139–145. doi:10.1023/B:EUPH.0000040511.46388.ef

Malosetti M, Ribaut J-M, van Eeuwijk FA (2013) The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. Front Physiol 4:1–17. doi:10.3389/fphys.2013.00044

Mathews K, Malosetti M, Chapman S, McIntyre L, Reynolds M, Shorter R, van Eeuwijk F (2008) Multi-environment QTL mixed models for drought stress adaptation in wheat. Theor Appl Genet 117(7):1077–1091. doi:10.1007/s00122-008-0846-8

Messina CD, Podlich D, Dong Z, Samples M, Cooper M (2011) Yield–trait performance landscapes: from theory to application in breeding maize for drought tolerance. J Exp Bot 62(3):855–868. doi:10.1093/jxb/erq329

Metzker ML (2010) Sequencing technologies – the next generation. Nat Rev Genet 11(1):31–46

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157(4):1819–1829

Muir W, Nyquist WE, Xu S (1992) Alternative partitioning of the genotype-by-environment interaction. Theor Appl Genet 84(1–2):193–200. doi:10.1007/bf00224000

Passioura JB (2012) Phenotyping for drought tolerance in grain crops: when is it useful to breeders? Funct Plant Biol 39(11):851–859. doi:10.1071/FP12079

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. PLoS Genet 2(12), e190. doi:10.1371/journal.pgen.0020190

Piepho HP (1998) Methods for comparing the yield stability of cropping systems. J Agron Crop Sci 180(4):193–213. doi:10.1111/j.1439-037X.1998.tb00526.x

Piepho H-P (2000) A mixed-model approach to mapping quantitative trait loci in barley on the basis of multiple environment data. Genetics 156(4):2043–2050

Piepho HP (2009) Ridge regression and extensions for genomewide selection in maize. Crop Sci 49(4):1165–1176

Piepho HP, Möhring J (2005) Best linear unbiased prediction of cultivar effects for subdivided target regions. Crop Sci 45(3):1151–1159. doi:10.2135/cropsci2004.0398

Piepho H, Möhring J, Melchinger A, Büchse A (2008) BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161(1):209–228. doi:10.1007/s10681-007-9449-8

Podlich DW, Cooper M, Basford KE, Geiger HH (1999) Computer simulation of a selection strategy to accommodate genotype environment interactions in a wheat recurrent selection programme. Plant Breed 118(1):17–28. doi:10.1046/j.1439-0523.1999.118001017.x

Poland JA, Rife TW (2012) Genotyping-by-sequencing for plant breeding and genetics. Plant Genome 5(3):92–102. doi:10.3835/plantgenome2012.05.0005

Poland J, Endelman J, Dawson J, Rutkoski J, Wu S, Manes Y, Dreisigacker S, Crossa J, Sánchez-Villeda H, Sorrells M, Jannink J-L (2012a) Genomic selection in wheat breeding using genotyping-by-sequencing. Plant Genome 5(3):103–113. doi:10.3835/plantgenome2012.06.0006

Poland JA, Brown PJ, Sorrells ME, Jannink J-L (2012b) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One 7(2), e32253. doi:10.1371/journal.pone.0032253

Prasanna B, Araus J, Crossa J, Cairns J, Palacios N, Das B, Magorokosho C (2013) High-throughput and precision phenotyping for cereal breeding programs. In: Gupta PK, Varshney RK (eds) Cereal genomics II. pp 341–374. doi:10.1007/978-94-007-6401-9_13

Reymond M, Muller B, Leonardi A, Charcosset A, Tardieu F (2003) Combining quantitative trait loci analysis and an ecophysiological model to analyze the genetic variability of the responses of maize leaf growth to temperature and water deficit. Plant Physiol 131(2):664–675. doi:10.1104/pp. 013839

Reymond M, Muller B, Tardieu F (2004) Dealing with the genotype×environment interaction via a modelling approach: a comparison of QTLs of maize leaf length or width with QTLs of model parameters. J Exp Bot 55(407):2461–2472. doi:10.1093/jxb/erh200

Reynolds M, Foulkes MJ, Slafer GA, Berry P, Parry MAJ, Snape JW, Angus WJ (2009a) Raising yield potential in wheat. J Exp Bot 60(7):1899–1918. doi:10.1093/jxb/erp016

Reynolds M, Manes Y, Izanloo A, Langridge P (2009b) Phenotyping approaches for physiological breeding and gene discovery in wheat. Ann App Biol 155(3):309–320. doi:10.1111/j.1744-7348.2009.00351.x

Riedelsheimer C, Endelman JB, Stange M, Sorrells ME, Jannink J-L, Melchinger AE (2013) Genomic predictability of interconnected biparental maize populations. Genetics 194(2):493–503. doi:10.1534/genetics.113.150227

Romagosa I, Fox P (1993) Genotype × environment interaction and adaptation. In: Hayward MD, Bosemark NO, Romagosa I, Cerezo M (eds) Plant breeding: principles and prospects. Springer, Dordrecht, pp 373–390

Romagosa I, Borràs-Gelonch G, Slafer G, Eeuwijk F (2013) Genotype by environment interaction and adaptation. In: Savin R, Costa-Pierce B, Misztal I, Whitelaw CB, Christou P (eds) Sustainable food production. Springer, New York, pp 846–870. doi:10.1007/978-1-4614-5797-8_199

Sadras VO, Lawson C (2011) Genetic gain in yield and associated changes in phenotype, trait plasticity and competitive ability of South Australian wheat varieties released between 1958 and 2007. Crop Pasture Sci 62(7):533–549. doi:10.1071/CP11060

Sadras VO, Rebetzke GJ, Edmeades GO (2013) The phenotype and the components of phenotypic variance of crop traits. Field Crop Res 154:255–259. doi:10.1016/j.fcr.2013.10.001

Sarkar S (1999) From the reaktionsnorm to the adaptive norm: the norm of reaction, 1909–1960. Biol Philos 14(2):235–252. doi:10.1023/a:1006690502648

Schulz-Streeck T, Ogutu JO, Karaman Z, Knaak C, Piepho HP (2012) Genomic selection using multiple populations. Crop Sci 52(6):2453–2461. doi:10.2135/cropsci2012.03.0160

Slafer GA (2003) Genetic basis of yield as viewed from a crop physiologist's perspective. Ann Appl Biol 142(2):117–128. doi:10.1111/j.1744-7348.2003.tb00237.x

Slafer G, Kernich G (1996) Have changes in yield (1900–1992) been accompanied by a decreased yield stability in Australian cereal production? Aust J Agric Res 47(3):323–334. doi:10.1071/AR9960323

Slafer G, Rawson H (1994) Sensitivity of wheat phasic development to major environmental factors: a re-examination of some assumptions made by physiologists and modellers. Funct Plant Biol 21(4):393–426. doi:10.1071/PP9940393

Smith A, Cullis B, Thompson R (2001a) Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. Biometrics 57(4):1138–1147. doi:10.1111/j.0006-341X.2001.01138.x

Smith AB, Cullis BR, Appels R, Campbell AW, Cornish GB, Martin D, Allen HM (2001b) The statistical analysis of quality traits in plant improvement programs with application to the mapping of milling yield in wheat. Aust J Agric Res 52(12):1207–1219. doi:10.1071/AR01058

Smith AB, Cullis BR, Thomson R (2005) The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. J Agric Sci 143(06):449–462. doi:10.1017/S0021859605005587

Smith AB, Lim P, Cullis BR (2006) The design and analysis of multi-phase plant breeding experiments. J Agric Sci 144(05):393–409. doi:10.1017/S0021859606006319

Snape JW, Butterworth K, Whitechurch E, Worland AJ (2001) Waiting for fine times: genetics of flowering time in wheat. Euphytica 119(1–2):185–190. doi:10.1023/a:1017594422176

Spindel J, Wright M, Chen C, Cobb J, Gage J, Harrington S, Lorieux M, Ahmadi N, McCouch S (2013) Bridging the genotyping gap: using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations. Theor Appl Genet 126(11):2699–2716. doi:10.1007/s00122-013-2166-x

Stephens M (2013) A unified framework for association analysis with multiple related phenotypes. PLoS One 8(7), e65245. doi:10.1371/journal.pone.0065245

Tardieu F (2003) Virtual plants: modelling as a tool for the genomics of tolerance to water deficit. Trends Plant Sci 8(1):9–14. doi:10.1016/S1360-1385(02)00008-0

Tardieu F, Reymond M, Muller B, Granier C, Simonneau T, Sadok W, Welcker C (2005) Linking physiological and genetic analyses of the control of leaf growth under changing environmental conditions. Aust J Agric Res 56(9):937–946. doi:10.1071/AR05156

Uitdewilligen JGAML, Wolters A-MA, D'hoop BB, Borm TJA, Visser RGF, van Eck HJ (2013) A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. PLoS One 8(5), e62355. doi:10.1371/journal.pone.0062355

Utz HF, Melchinger AE, Schön CC (2000) Bias and sampling error of the estimated proportion of genotypic variance explained by quantitative trait loci determined from experimental data in maize using cross validation and validation with independent samples. Genetics 154(4):1839–1849

van der Heijden G, Song Y, Horgan G, Polder G, Dieleman A, Bink M, Palloix A, van Eeuwijk F, Glasbey C (2012) SPICY: towards automated phenotyping of large pepper plants in the greenhouse. Funct Plant Biol 39(11):870–877. doi:10.1071/FP12019

van Eeuwijk F, Denis J, Kang M (1996) Incorporating additional information on genotypes and environments in models for two-way genotype by environment tables. In: Kang MS, Gauch HG Jr (eds) Genotype-by-environment interaction. Taylor & Francis Group, Boca Raton, Florida, pp 15–50

van Eeuwijk FA, Malosetti M, Yin X, Struik PC, Stam P (2005) Statistical models for genotype by environment data: from conventional ANOVA models to eco-physiological QTL models. Aust J Agric Res 56(9):883–894. doi:10.1071/AR05153

van Eeuwijk FA, Bink MCAM, Chenu K, Chapman SC (2010) Detection and use of QTL for complex traits in multiple environments. Curr Opin Plant Biol 13(2):193–205. doi:10.1016/j.pbi.2010.01.001

Varshney RK, Terauchi R, McCouch SR (2014) Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. PLoS Biol 12(6), e1001883. doi:10.1371/journal.pbio.1001883

Verbyla AP, Cullis BR (2012) Multivariate whole genome average interval mapping: QTL analysis for multiple traits and/or environments. Theor Appl Genet 125(5):933–953

Wang E, Robertson MJ, Hammer GL, Carberry PS, Holzworth D, Meinke H, Chapman SC, Hargreaves JNG, Huth NI, McLean G (2002) Development of a generic crop model template in the cropping system model APSIM. Eur J Agron 18(1–2):121–140. doi:10.1016/S1161-0301(02)00100-4

West BT, Welch KB, Galecki AT (2006) Linear mixed models: a practical guide using statistical software. Boca Raton, Florida

Windhausen VS, Atlin GN, Hickey JM, Crossa J, Jannink J-L, Sorrells ME, Raman B, Cairns JE, Tarekegne A, Semagn K, Beyene Y, Grudloyma P, Technow F, Riedelsheimer C, Melchinger AE (2012) Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. G3 Genes Genomes Genet 2(11):1427–1436. doi:10.1534/g3.112.003699

Woltereck R (1909) Weitere experimentelle Untersuchungen über Artveränderung, speziel über das Wesen quantitativer Artunterschiede bei Daphnien. Verhandlungen der deutschen zoologischen Gesellschaft 19:110–173

Wright S (1931) Evolution in mendelian populations. Genetics 16(2):97

Wright S (1932) The roles of mutation, inbreeding, crossbreeding and selection in evolution. In: Proceedings of the 6th international congress of genetics, Brooklin, NY, pp 356–366

Yan W, Hunt LA, Sheng Q, Szlavnics Z (2000) Cultivar evaluation and mega-environment investigation based on the GGE biplot. Crop Sci 40(3):597–605. doi:10.2135/cropsci2000.403597x

Yin X, Chasalow SD, Dourleijn CJ, Stam P, Kropff MJ (2000) Coupling estimated effects of QTLs for physiological traits to a crop growth model: predicting yield variation among recombinant inbred lines in barley. Heredity 85(6):539–549. doi:10.1046/j.1365-2540.2000.00790.x

Yin X, Stam P, Kropff MJ, Schapendonk AHCM (2003) Crop modeling QTL mapping, and their complementary role in plant breeding. Agron J 95(1):90–98. doi:10.2134/agronj2003.9000

Yin X, Struik PC, Kropff MJ (2004) Role of crop physiology in predicting gene-to-phenotype relationships. Trends Plant Sci 9(9):426–432. doi:10.1016/j.tplants.2004.07.007

Zhao Y, Gowda M, Longin F, Würschum T, Ranc N, Reif J (2012) Impact of selective genotyping in the training population on accuracy and bias of genomic selection. Theor Appl Genet 125(4):707–713. doi:10.1007/s00122-012-1862-2

Zheng B, Chenu K, Fernanda Dreccer M, Chapman SC (2012) Breeding for the future: what are the potential impacts of future frost and heat events on sowing and flowering time requirements for Australian bread wheat (Triticum aestivum) varieties? Glob Chang Biol 18(9):2899–2914. doi:10.1111/j.1365-2486.2012.02724.x

Zheng B, Biddulph B, Li D, Kuchel H, Chapman S (2013) Quantification of the effects of VRN1 and Ppd-D1 to predict spring wheat (Triticum aestivum) heading time across diverse environments. J Exp Bot 64(12):3747–3761. doi:10.1093/jxb/ert209

Chapter 4
Process-Based Simulation Models Are Essential Tools for Virtual Profiling and Design of Ideotypes: Example of Fruit and Root

Michel Génard, Mohamed-Mahmoud Memmah, Bénédicte Quilot-Turion, Gilles Vercambre, Valentina Baldazzi, Jacques Le Bot, Nadia Bertin, Hélène Gautier, Françoise Lescourret, and Loïc Pagès

Abstract Process-based simulation models (PBSMs) combine, in various mathematical frameworks, many biological functional hypotheses on responses of plant processes to environmental fluctuations. Model simulated responses can be analysed in the context of adapting the current agricultural systems to the changing environment. From loads of simulations made with various cultural practices, these models allow the virtual profiling of plants and a mere analysis of how processes interact when crops are perturbed by one or several changes. They allow also describing the development of plant traits as a consequence of environmental and genetic conditions. Such knowledge is required to decipher the genotype×environment×management (G×E×M) interactions so as to build genotypes adapted to particular conditions, i.e., plant ideotypes. Two PBSMs dealing with (1) fruit quality and sensitivity to diseases and (2) root system architecture, respectively, are shortly described in this chapter. These models have been used to analyse various fruit and root properties, to deconvolute G×E×M interactions and to identify ecophysiological traits related to crop yield improvement, root foraging performance and fruit quality. PBSMs appear to be powerful tools to phenotype plants at the process level in a comprehensive and "costless" way.

M. Génard (✉) • M.-M. Memmah • B. Quilot-Turion • G. Vercambre • V. Baldazzi • J. Le Bot • N. Bertin • H. Gautier • F. Lescourret • L. Pagès
INRA, UR1115 Plantes et Systèmes de Culture Horticoles, 84000 Avignon, France
e-mail: michel.genard@paca.inra.fr

X. Yin, P.C. Struik (eds.), *Crop Systems Biology*,
DOI 10.1007/978-3-319-20562-5_4

4.1 Introduction

Plant production systems evolve regularly so as to cope with a growing worldwide demand of food production and important issues of quality, plant health, food safety and respect for the environment. Moreover, this adaptation must encompass the global climate change context. As a consequence, the most critical question for our future could be how to design the best possible combinations of genetic resources and environmental conditions (or cultural practices) able to achieve such objectives.

A prerequisite for performing such a design concerns the analysis of genotypic and environmental effects on plant physiological traits governing development and functioning and also on traits related to plant production, product quality, and plant health. Many plant scientists put the focus on genotype studies using profiling technologies. These techniques probe many genes, transcripts, proteins or metabolites, at once. The profiling technologies allow the analysis of system-wide responses (Hennig 2007; Kopka et al. 2004), which will undoubtedly help to understand 'how things work' at the cellular level. However, it is not always easy to interpret clearly the observed '*omic*' changes in terms of plant functioning because gene functions may be known at the cellular level but they are rarely identified in terms of plant responses while our knowledge of gene regulation in relation to the environment remains weak. In addition to these studies, one way to improve our understanding of plant responses is to use process-based simulation models (PBSMs). From the making of a load of simulation runs under various environmental conditions and/or cultural practices, it is possible to use these models to perform a virtual profiling (see below) and a mere analysis of how the plant system works, i.e., how the numerous processes interact when the plant is perturbed by one or several changes at the cellular level (Génard et al. 2010). Indeed, PBSMs offer a theory describing how the system components causally interact to produce a given outcome. From this viewpoint, Peck (2004) regards simulation as "the creation of a possible world that is constructed *in silico* using computer programs to represent the processes under consideration".

In line with this idea, some agronomists and geneticists have proposed a smart approach that consists of analysing variation in the development of plant traits via PBSMs as a consequence of environmental and genetic factors (Hammer et al. 2005; Yin et al. 2005). Such PBSMs allow to virtually analyse how plants react to changing environments, cultural practices and genetic variability. The main expectation of such an approach is to decipher genotype × environment × management (G × E × M) interactions (Asseng et al. 2002; Boote et al. 2003; Chapman et al. 2003; de Dorlodot et al. 2007) to build genotypes adapted to particular conditions, e.g., critical pedoclimatic situation, innovative crop management, future climates (Hammer et al. 2002). Thus, in contrast with the strategy developed in the past, researchers do not seek any longer genotypes adapted to all conditions but particular ones. Indeed, specifically adapted genotypes perform relatively better than other ones under a set

of conditions of particular interest, or lead to better environment-friendly production systems. Progress in this direction will clearly depend on the genetic information available on these related processes that will be injected in the PBSMs. This specific integration is the subject discussed in Chap. 1 of this book by Baldazzi et al.

New 'ideotypes' are real or virtual plant cultivars expressing an ideal phenotype adapted to target environments (Letort et al. 2008; Tardieu 2003). To design ideotypes, PBSMs are viewed as essential tools. Indeed, within a target environment, they can simulate a large number of virtual genotypes, each one being characterized by a set of genetic parameters. The challenge is now to identify a few of these 'ideotypes' among a myriad of simulated genotypes. The first attempts were conducted using techniques such as trial and error methods (Haverkort and Grashoff 2004; Herndl et al. 2007) and were quickly confronted with the difficulty and the hardness of the task. This is especially the case when multi-objectives are targeted. Indeed, the design of innovative cultivars is based on strong nonlinear antagonistic criteria with respect to influential constraints of production or environment. Resolving such problems is difficult using classical methods and it is known to be a matter of nonlinear multi-objective optimisation. Thanks to collaborations between biologists and mathematicians, effective methods have been proposed recently. They emerged from the field of multi-objective optimisation algorithms, e.g., genetic (Letort et al. 2008) and particle swarm (Qi et al. 2010) algorithms.

Two PBSMs dealing with (1) fruit quality and sensitivity to diseases and (2) root system architecture, respectively, will be shortly described in this chapter and used to analyse various fruit and root properties, sometimes including what Trewavas (2003) called *memory* and *compensatory* effects. The strength of PBSMs for analysing crop systems and for performing virtual profiling will be illustrated. We will also show how the use of model-based sensitivity analysis serves the selection of genetic traits necessary to design ideotypes. Finally, an approach for designing ideotypes will be described using the "fruit quality and sensitivity to diseases" PBSM.

4.2 What Are Process-Based Simulation Models for Fruit and Root Systems?

Process-based simulation models are collections of hypotheses and rules about the interrelationships linking processes to environmental variations and producing responses that can be analysed. The results are generally produced in the time domain and at an appropriate time step of the studied processes. They provide, therefore, a basis for the understanding of developmental, physiological and genetic phenomena, by dissecting complex traits into "elementary" processes. The classical notion of a single limiting factor is replaced by the idea of a sequence and/or network of different limitations operating through the plant's lifecycle. These interconnections and feedback regulations among the system components generate

unexpected global system properties, called emergent properties, which do not appear when the subsystems are individually considered (Trewavas 2006). Genotype × Environment (G × E) interactions are emergent properties of the whole system in which several processes interact. However, these interactions can also operate at the process level.

PBSMs permit the quantification of plant or organ responses to genetic, environmental, and management factors within a mathematical framework in which parameters are genotype-specific, thus allowing dynamic simulations of biophysical and physiological processes. They have yet been successfully used to deconvolute G × E interactions and to identify ecophysiological traits in studies designed to improve crop yield (Yin et al. 2000), root foraging performance (Pagès 2011), phenological development (Stewart et al. 2003; Welch et al. 2005), leaf elongation rate (Reymond et al. 2003) and fruit quality (Quilot et al. 2005).

Fruit quality traits have been seldom subjected to modelling, probably because they result from a poorly understood chain of processes encompassing only partly known steps of the complex underlying mechanisms (Struik et al. 2005). Conversely, root system architecture has been modelled for the last two decades (Dunbabin et al. 2013) but the models are generally too complex to give rise to ideotype conception (Pagès et al. 2004). PBSMs focusing on fruit size and composition have been proposed recently (Génard et al. 2007). They have proven their ability to produce emergent properties, i.e., to handle perturbations to any process and self-correct them as real plants do. The structure and some properties of these models will be presented hereafter.

4.2.1 Virtual Fruit Model

The first attempt towards integration and simulation of multiple quality traits has been made in the Virtual Peach Fruit model (Lescourret and Génard 2005). This PBSM (Fig. 4.1a) integrates many sub-models dealing with fruit growth and quality elaboration, and had its genesis in a model proposed by Lescourret and Génard (2005) including three existing process-based sub-models describing dry mass, sugar and water accumulation in the fruit flesh. Then, supplementary sub-models were added to describe skin conductance and microcracking (Gibert et al. 2005), respiration and citric acid accumulation (Lobit et al. 2003; Wu et al. 2007), and ethylene emission (Génard and Gouble 2005).

For now, the Virtual Peach fruit model describes the carbon (C) balance of a fruit-bearing stem. The available daily pool of C assimilates builds up from leaf assimilation plus possible C mobilized from the reserves. Carbon is allocated according to organ demands and priority rules. The fruit flesh is assumed to behave as one big cell. The carbon flow entering the fruit is partitioned into several compounds: four sugars (sucrose, sorbitol, glucose and fructose), other fruit compounds globally considered and the respired CO_2. Water flows into the fruit following the

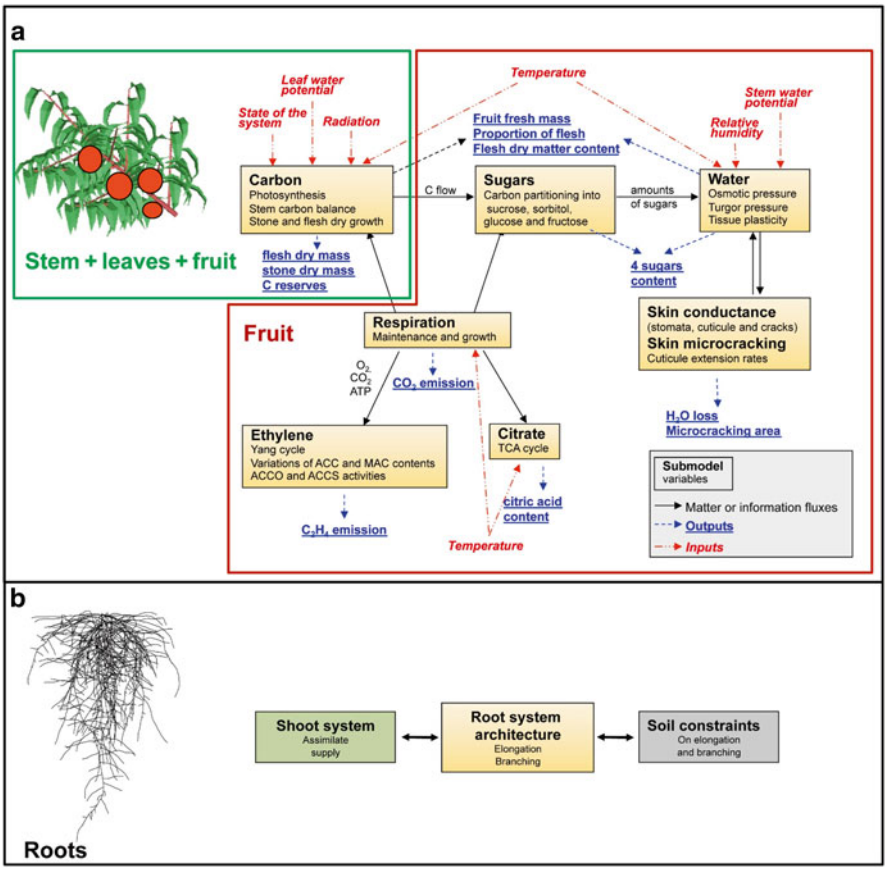


Fig. 4.1 Virtual fruit and root system architecture models. (**a**) Schematic representation of the relationships between sub-models as considered in the Virtual Fruit model. The sub-models simulate carbon balance of a fruit bearing stem, sugars and citric acid metabolism within the fruit, fruit water balance, skin conductance and microcracking, fruit respiration and ethylene metabolism. The inputs of the model are weather data, leaf and stem water potential, and the number of leafy shoots and fruits on the stem. The outputs (*underlined*) are flesh and stone masses, sugar and acid contents, skin microcracking and emission of gases. (**b**) The root system architecture model combines two main developmental processes (root elongation and branching) which are described by a number of rules with parameters. These processes are also affected by the assimilate supply (from the shoot system) and by the soil constraints

differences of hydrostatic and osmotic pressures between the xylem or phloem and the fruit. Changes in the fruit volume are predicted by the Lockhart equation as a function of turgor pressure. Fruit transpiration is calculated from the overall skin conductance to water vapour, by adding the individual conductance of stomata, cuticle and cracks. Cracks are assumed to happen when the pulp expansion rate exceeds that of the cuticle. The rate of citrate metabolism is calculated as the product

of a 'synthesis potential' by an 'efficiency level', which depends on respiration intensity. The ethylene sub-model simulates its biosynthetic pathway as a function of ATP supply, O_2 and CO_2 tissue concentrations, and its regulation by 1-aminocyclopropane-1-carboxylic acid (ACC) synthase and ACC oxidase. The inputs of this PBSM are weather data (i.e., global radiation, temperature and air relative humidity), leaf and stem water potential, and the number of leafy shoots and fruits on the stem. The parameters of the model have been estimated for the peach species in the framework of a large research programme undertaken over the last 10 years.

The Virtual Peach Fruit model has allowed the simulation of complex behaviours of fruit growth and quality traits in response to environmental fluctuations (Fig. 4.2). For example, compared with optimal water supply (W), the response to the alternation of periods of restriction and normal irrigation (S/W) shows a "compensatory growth" phenomenon, observable after re-watering when fruits regain that bit of growth lost during the period of stress. The model simulations show also clearly that fruits facing continuous water stress (S) maintain their growth rate, while fruits becoming water-stressed after a period of normal watering (W/S) experience a sharp slowdown in growth. This implies that S fruits adapted to drought but W/S fruits did not. In the model, these growth patterns are related to sugar concentration changes, the sustained fruit growth being linked with high sugar concentrations under conditions of water deficit. In real plants, Trewavas (2004) called this kind of adaptation a 'memory effect'. He reckons that compensatory growth is a corrective mechanism involving a feedback control to achieve a developmental goal (Trewavas 2003). We can hypothesize that the Virtual Peach fruit model mimics such a "sugar signal"-based mechanism, if we assume that the increase in sugar concentration during the stress period (as shown in Fig. 4.2) promotes growth after re-watering.

These responses are not accounted for by any of the Virtual Fruit sub-models taken separately, but they result from feedback regulations, which emerge from the entire Virtual Fruit model. This illustrates the usefulness of PBSM to analyse complex responses of quality traits to environmental variations.

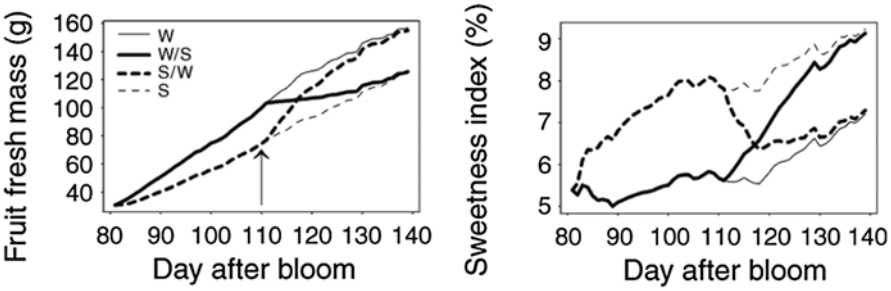


Fig. 4.2 Time course of fruit fresh mass and sweetness (correlated to sugar concentration) simulated by the Virtual Peach Fruit model under four scenarios of water conditions (W = normal water condition, W/S = normal then stressed water condition, S/W = stressed then normal water condition, S = water stress). The *arrow* indicates the time when the water condition changed

4.2.2 Root System Architecture Model

Pagès (2011) presented the main aspects of a model of root system architecture (RSA) to link elementary developmental processes (and associated traits) at the individual root level to complex (or integrated) traits describing foraging performance at the whole root system level. This model was then calibrated and evaluated on a panel of different species belonging to various plant families (Pagès et al. 2014). It is a discrete model simulating the 3D RSA with a 1-day time step. The dynamic virtual root system is represented as a set of small segments with different attributes (location, diameter, age, connection). Basically, the model includes three types of interacting processes: (*i*) a number of morphogenetic rules define the elongation and branching of individual roots; (*ii*) the overall demand of roots for growth is compared with the biomass allocation to the root system within the whole plant and can be reduced in the case of limitation; (*iii*) each root meristem (located in the soil space) experiences local conditions, which can reduce its potential elongation and branching (e.g., strong soil or fresh temperatures).

Regarding the morphogenetic rules governing the RSA dynamics, it is noticeable that the size of the root meristems plays a central role. Potential elongation rates of individual roots are linearly linked to their tip diameter, the slope of this relationship being assumed to be a genetic parameter. For each species, the elongating-roots' tip diameter ranges between the minimal and maximal values that are both specified as genetic parameters. The branching process defines both the longitudinal spacing of successive lateral roots on their mother, and the tip diameter distribution among lateral roots. On average, their diameter is linearly linked to that of their mother root and it is supposed to be variable, the coefficient of variation being also a genetic parameter.

This very simple model (ArchiSimple) benefited from a long-term experience on RSA modelling (Pagès and Ariès 1988), which facilitated the necessary simplification (Fig. 4.1b). However, it was designed to include ecophysiological concepts such as sink strength (through the meristem size), possible carbon source limitations and the effects of some soil characteristics to modulate root elongation (e.g., soil strength or soil temperature). These processes are known to exhibit strong genetic variations (de Dorlodot et al. 2007).

From the virtual RSA, it was possible to estimate a foraging performance (Pagès 2011), considering that root systems take up soil nutrients located around the roots, within a given rhizosphere volume characterised by its distance to the roots. This distance reflects in fact the type of nutrient that is targeted. For non-mobile ions (e.g., phosphate) this distance is only a few millimetres whereas for mobile ions (e.g., nitrate) or water it attains several centimetres. It is worth noticing that an integrated trait such as the colonised root volume can be calculated on virtual 3D root systems, but it is actually impossible to assess on real root systems.

Thus, in this case, the model, that can be used to design root ideotypes, includes in fact an RSA model explicitly connected to the soil and the shoot system and a very simplified uptake model representing the mere limiting soil resources (water and/or nutrients).

4.3 Virtual Profiling on Roots and Fruits

Our purpose is to show how PBSMs can be used to perform virtual profiling to phenotype plants in terms of ecophysiological processes. The following examples describe two possible, but not unique, approaches. On the root systems, our profiling proceeded from the analysis of a few parameters studied on a large virtual root population while on fruits, it came from the study of a large set of variables on two contrasted genotypes.

4.3.1 Roots

Using the ArchiSimple model, a large number of root systems were simulated and used to explore the relationships between their overall performance and their elementary developmental traits.

For this purpose, each parameter was given random values in its plausible interval, defined from the literature. The experimental design conformed to a paving, as suggested by Saltelli et al. (2008). Root foraging efficiency was defined as the ratio of the total rhizospheric volume to a theoretical volume without any root overlap, i.e., with no redundant soil exploration. From this in silico experimentation several main results were obtained, which illustrates the interest of such approaches. The overall foraging efficiency was highly dependent on the overall size of the root system, because the rhizosphere of different roots tended to overlap more and more during root system development and extension (Fig. 4.3). As expected, this overlap was dependent on the rhizosphere size (distance). The approach allowed quantifying this important phenomenon, which is usually neglected, even though root systems are particular objects in which roots are aggregated in space, mainly because they are connected as a branching system. The efficiency varied considerably from one root system to another, even for a same total root length, and the variations were highly dependent on some parameters. Favourable ranges were defined for particular parameters, such as the inter-branch distance. Moreover, the overall tip-diameter range allowed by the virtual phenotype (called heterorhizy) was shown favouring efficiency. Interestingly, the genetic parameter modulating the variation among lateral root diameters had also a large impact on the overall efficiency. When a subpopulation of efficient root systems was selected (elite population), it was shown that some parameter associations were excluded and conversely other associations between parameter values were favoured. Thus, correlations between parameters enabled to quantify the trade-offs in the elite population.

Thus overall, the use of such a PBSM confirmed its value to decipher these very complex relationships and allowed to bridge two scales: one on which developmental processes can be studied, and the other on which agronomic performances can be evaluated.

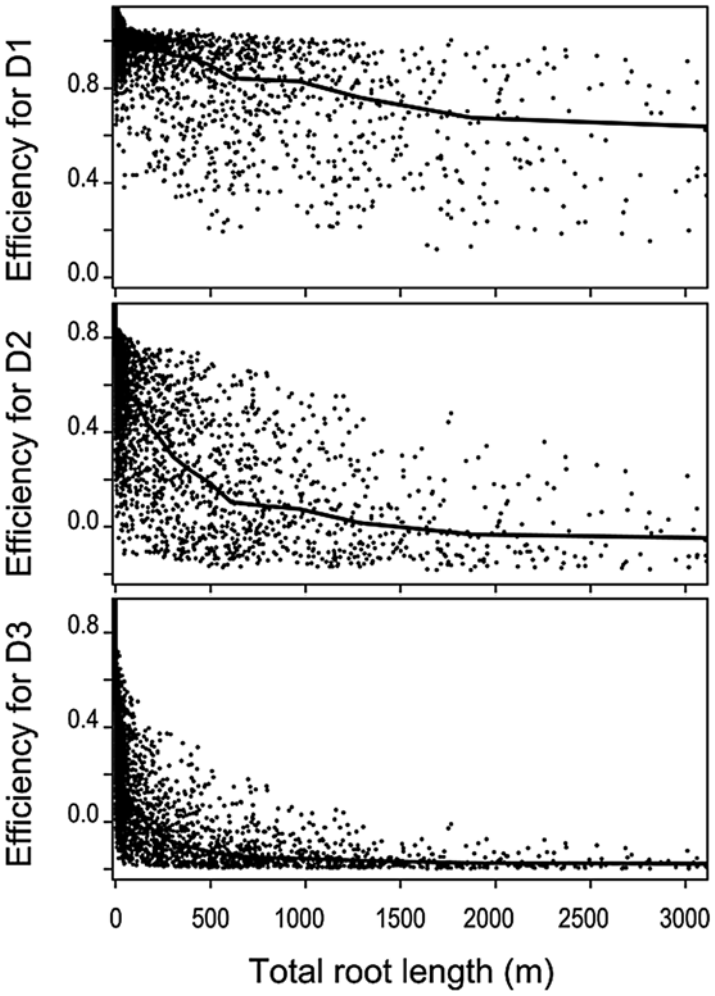


Fig. 4.3 Relationship between foraging efficiency (quantifying the rhizosphere overlap between different roots) and size of the root system (redrawn from Pagès (2011)). Efficiency varies between 0 (full overlap) and 1 (no overlap). The considered rhizosphere radius was 3 mm (D1), 10 mm (D2) or 30 mm (D3). Each point represents a root system simulated by the model of Pagès (2011) on which the total length and the rhizosphere characteristics have been calculated. The *lines* represent the trends, calculated with the "loess" function in the R software

4.3.2 Fruit

The 'Virtual Fruit' model has been used to simulate wild type and mutant plants over a period of 70 days for fruit growth. To simulate the mutation in the model, we decreased by 70 % the value of the unique parameter that modulates the fruit sugar uptake. We used the model to calculate the daily value of 39 variables related to

various processes (Table 4.1). The considered functional variables were in most cases rates such as photosynthetic rate (g CO_2 m^{-2} d^{-1}) or relative rates such as daily variation of sucrose content (d^{-1}). Some physical characteristics such as turgor pressure or conductance were also considered since they are proportional to relative growth rate or fluxes.

By analogy with some current presentation of 'omic' results, we produced a heat map of the 39 variables (Fig. 4.4), which appears as a virtual profile fingerprint summarizing the processes impacted by fruit ageing. The wild-type genotype showed very contrasted temporal variation patterns according to the variables. Indeed, the relative variation of ATP, glucose and fructose suffered large temporal fluctuation mainly due to changes in the climatic environment. Other variables such as fruit turgor pressure showed slow oscillations during fruit development. Most of the variables followed a temporal gradient with three distinct periods. Early in the season there was a high production of citric acid in the fruit. During the mid-season we detected high activities of stone growth and sucrose accumulation (SU). In the late period of fruit maturation, ethylene production increased.

For the mutant plants, the general pattern was similar to that observed for the wild-type genotype. The same variables were involved in three main growth periods and oscillatory behaviours were also simulated (Fig. 4.4). During the first period, the variables involved remained almost quantitatively unchanged. By contrast, changes appeared in the two last periods since the mutation triggered a large effect in several leaf and fruit variables (photosynthesis, growth, respiration and metabolism) and delayed the fruit developmental rate. All variables involved in the mid-season period had lower intensities compared with the wild-type genotype. Similarly, most variables involved in the fruit maturation period showed lower activities, except enzymes involved in the sucrose and sorbitol metabolisms (Eso and Esu), which maintained similar activities. The flesh osmotic potential (Os) was the unique oscillatory variable showing higher values in the mutant.

At the end, the mutation involved in the fruit sugar uptake had a strong effect in most fruit processes and on the plant source activity throughout fruit growth. The single mutation perturbed the whole system and impacted fruit quality, decreasing fruit mass and skin cracking and lowering ethylene emission (Fig. 4.5). Similar results were also obtained with a change in fruit load, which impacted also most of the plant and fruit processes (Génard et al. 2010).

Such a virtual profiling approach could lead to new ways of exploring *in silico* the impact of mutations or naturally occurring genetic variations. A thrilling challenge for the future will be to connect virtual and 'omic' profilings. Two main approaches are possible, one being empirical and the other mechanistic. A means to facilitate virtual and 'omic' profiling connection is to perform 'omic' and virtual profiling on the same subjects and to use data-mining technologies such as the bidirectional orthogonal projection to latent structures (O2-PLS) method recently proposed by Bylesjö et al. (2007) to search for links between them. To go further, the mechanistic integration of information generated by 'omics' technologies into

Table 4.1 List of the processes and functional variables of the Virtual Fruit model considered for the virtual profiling. Label, unit and range of variation are given for each variable

Organ	Process	Variable	Label	Unit	Range of values (in Fig. 4.4)
Stem and Shoots	Carbon balance	Relative stem C-storage rate	Cst	d^{-1}	−0.0022 to 0
		Relative shoot growth rate	Sh	d^{-1}	0 to 0.1
		Relative shoot C-storage rate	Csh	d^{-1}	−0.014 to 0.009
		Relative stem and shoot maintenance respiration	MRs	d^{-1}	0.0009 to 0.0017
		Leaf photosynthetic rate	Lph	$g\ m^{-2}\ d^{-1}$	1.37 to 4.01
		Maximal leaf photosynthetic rate	Pm	$\mu mol\ CO_2\ m^{-2}\ s^{-1}$	5.54 to 19.49
Fruit	Carbon balance	Relative fruit demand	Fde	d^{-1}	0.0012 to 0.0318
		Fruit photosynthetic rate	Fph	d^{-1}	0.004 to 0.187
		Relative flesh respiration rate	Fre	$\mu mol\ CO_2\ g^{-1}\ DM\ d^{-1}$	0.077 to 0.348
		Relative flesh dry mass growth rate	Fd	d^{-1}	0.0031 to 0.0459
		Relative stone dry mass growth rate	Sd	d^{-1}	0.0019 to 0.0244
	Water balance	Relative transpiration	Tr	d^{-1}	0.004 to 0.104
		Relative water influx	Flux	d^{-1}	0.07 to 0.36
		Flesh turgor pressure	P	bar	5.0 to 5.15
		Flesh osmotic potential	Os	bar	11.6 to 14.6
		Relative flesh fresh mass growth rate	Ff	d^{-1}	0.02 to 0.36
		Relative stone fresh mass growth rate	Sf	d^{-1}	0.0013 to 0.0226
	Skin	Crack conductance	Crco	$cm\ h^{-1}$	0.0 to 81.2
		Stomatal conductance	Sco	$cm\ h^{-1}$	11.5 to 27.7
		Cuticular conductance	Cuco	$cm\ h^{-1}$	59.8 to 357.1
		Skin area growth rate	Skin	$cm^2\ d^{-1}$	0.07 to 1.64

(continued)

Table 4.1 (continued)

Organ	Process	Variable	Label	Unit	Range of values (in Fig. 4.4)
	Sugar and citric acid metabolism	Relative daily sucrose content variation	SU	d^{-1}	−0.0007 to 0.0055
		Relative daily sorbitol content variation	SO	d^{-1}	−0.0027 to 0.0039
		Relative daily glucose content variation	GL	d^{-1}	−0.0009 to 0.0022
		Relative daily fructose content variation	FR	d^{-1}	−0.0004 to 0.0022
		Metabolizing sucrose enzyme activity	Esu	d^{-1}	0.0008 to 0.041
		Metabolizing sorbitol enzyme activity	Eso	d^{-1}	0.32 to 0.89
		Activity of enzyme metabolizing glucose and fructose in other compounds	Eglfr1	d^{-1}	0.015 to 0.218
		Activity of enzyme metabolizing glucose and fructose for respiration	Eglfr2	d^{-1}	1.15 to 9.19
		Relative daily citrate content variation	CI	d^{-1}	−0.71 to 2.19
	Ethylene metabolism	Relative daily ACC content variation	ACC	mol g^{-1} d^{-1}	−0.008 to 0.009
		Relative daily MACC content variation	MACC	mol g^{-1} d^{-1}	−0.0008 to 0.0031
		Relative daily C_2H_4 content variation	C2H4	mol g^{-1} d^{-1}	−0.0003 to 0.0003
		Relative daily O_2 content variation	O2	mol g^{-1} d^{-1}	−0.051 to 0.075
		Relative daily CO_2 content variation	CO2	mol g^{-1} d^{-1}	−0.042 to 0.029
		Relative daily ATP content variation	ATP	mol g^{-1} d^{-1}	−0.029 to 0.022
		C_2H_4 emission	C2H4em	µmol g^{-1} h^{-1}	0 to 0.001
		ACCO activity	ACCo	µmol g^{-1} h^{-1}	0 to 0.0007
		ACCS activity	ACCs	µmol g^{-1} h^{-1}	0 to 0.0025

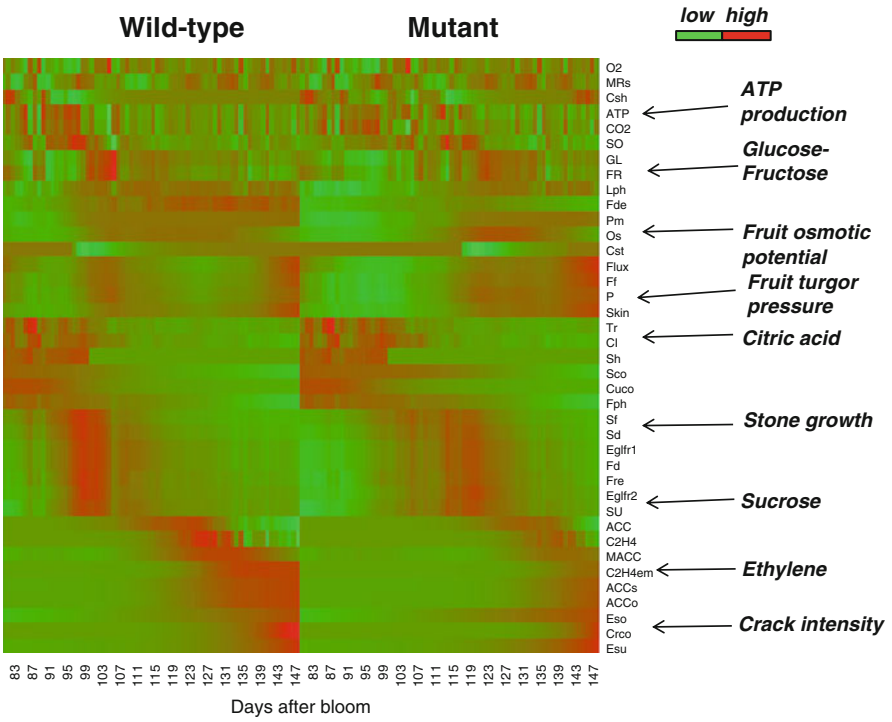


Fig. 4.4 Heat map surface of model variables during the fruit developmental period of the wild type and mutant genotypes. In the heat map, each row is a variable and each column is a day after full bloom. Abbreviations of the variables are defined in Table 4.1. The values increase from *green* to *red*

models could provide a global view of how plants operate. So far, upscaling attempts from gene to cell have been successfully undertaken (Tomita et al. 1999), but further upscaling towards the organ or the plant levels has not yet been performed (Baldazzi et al. 2012). A possible avenue would be to embed mechanistic metabolic models into PBSMs such as the Virtual Fruit model. For instance it could be possible to substitute its current simplified sugar model with a more detailed one, such as the sugar metabolism model developed by Uys et al. (2007) for sugarcane. The Virtual Fruit model would thus produce inputs for the metabolic model, which could in return simulate the production of metabolites along the metabolic pathways. By comparing simulations with real metabolic profiles obtained through metabolomic studies, it would be possible to test the benefit of inserting such metabolic model within the Virtual Fruit model.

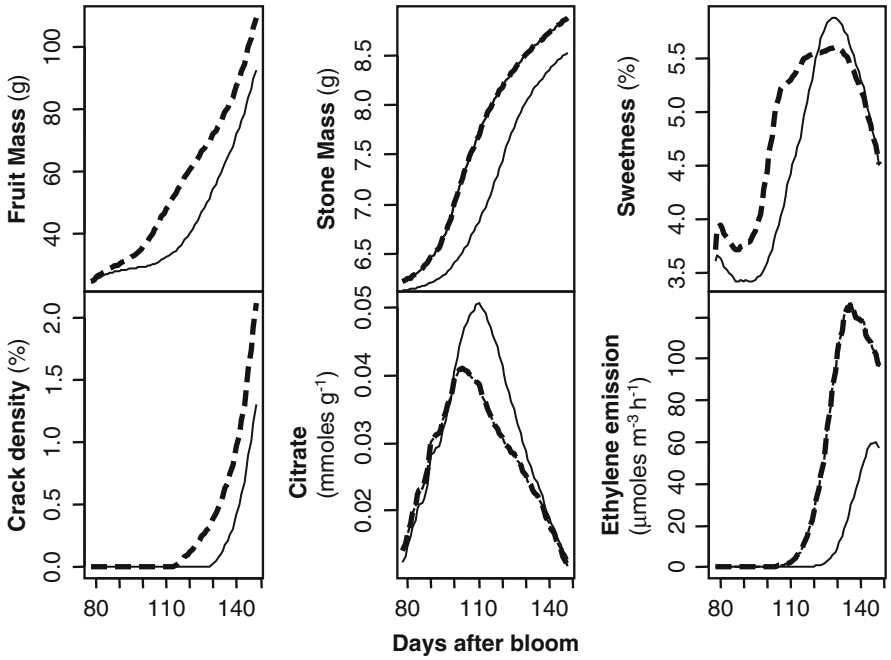


Fig. 4.5 Comparison of wild-type (*thick dashed lines*) and mutant (*thin lines*) simulation outputs generated by the Virtual Fruit model. Variation of fruit growth, sweetness (equivalent sucrose), citric acid content, cracking and ethylene production are presented during the fruit developmental period

4.4 Ideotype Design

4.4.1 Sensitivity Analysis, a Key-Step Before Designing Ideotypes

Sensitivity analysis (SA) is a statistical technique allowing to assess the impact of changing some input parameters on the model outputs (Blower and Dowlatabadi 1994). Saltelli et al. (2008) suggested conducting SA on the model (i) to test its accuracy, (ii) to prioritize the parameters before their estimation, (iii) to simplify the model and reduce its parameter number, and (iv) to identify the interactions between parameters. For example, ranking model parameters arises from sensitivity indices reflecting the main effect of each parameter and their interactions. Parameters having small or no effect on model outputs can thus be set to fixed values, leading to model simplification.

We can distinguish local SA methods from global ones. The former evaluate the impact of a very small variation around a given input value, while the latter study the output variability when that of the input covers the whole possible domain (Jacques 2011). Many SA methods have been used in the literature. These methods

have different bases, including elementary effects (Morris or *one-step-at-a-time* method), regression and correlation coefficients, variance decomposition (Sobol, FAST). Software and packages implementing these methods are largely accessible, for instance the *R sensitivity package*.[1] SA techniques are very useful to study the behaviour of complex numerical models such as ecophysiological models described in this chapter. Therefore, SA techniques were among the earliest tools used for the model-based design of ideotypes (Habekotté 1997). Indeed, SA methods coupled with process-driven biophysical models may help answering some '*what if*' questions before engaging in experiments (Fischer 1996). A SA of ecophysiological models under contrasted climatic conditions and/or agricultural practices allows identifying the most important parameters that mainly affect outputs of interest (i.e., targeted traits). For example, Quilot-Turion et al. (2012) performed an SA on the 'Virtual Fruit' model to identify the main parameters affecting fruit fresh mass, sweetness and crack density. The *'elementary effects'* screening method (Morris 1991; Saltelli et al. 1999) was used for this purpose. This method computes two sensitivity measures: the mean and the standard deviation of the distribution of the elementary effects associated with a given parameter. They respectively assess the overall influence of a given parameter on the output and on the interactions of a parameter with another one. In the study of Quilot-Turion et al. (2012), the sensitivity of each output variable to each parameter was quantified by considering a 10 % variation interval around each parameter reference value. For the 'Virtual Fruit' model, from about 60 parameters submitted to the SA, only few parameters had significant impact on the model outputs. We will show in the next section how these six genetic parameters can be combined to create interesting ideotypes.

Pagès et al. (2012) used a global SA on the root system architecture to link developmental parameters with the shape of the vertical root length distribution, which is often used to characterize the root system in crop models. This analysis allowed creating a meta-model that enabled model inversion to facilitate the estimation of developmental parameters from vertical root length density profiles (Pagès et al. 2012).

4.4.2 Multi-objective Optimisation Algorithms to Design Ideoypes

Designing environment-friendly production systems that produce safe food of good quality is an important challenge for the future. Indeed, such a production system may sustain the economic viability of farms. Today, one of the promising ways to tackle this issue is to identify the best combinations of genetic resources and cultural practices adapted to target environments. The idea here is to take advantage of the strong genotype×environment×management (G×E×M) interactions in

[1] http://cran.r-project.org/web/packages/sensitivity/index.html

order to design plant ideotypes that meet these objectives. Such an approach relies on two main pillars: the potentialities offered by integrating genetic information into process-based models and the merger of breeders and agronomists (Messina et al. 2009). Therefore, model-based design of ideotypes aims at finding the most suitable combinations of genetic parameters, genotype fingerprints and cultural practices adapted to target environments (Letort et al. 2008; Tardieu 2003).

It should be clearly stressed here that model-based ideotype design involves many conflicting criteria subject to hard constraints and therefore, it shows a number of difficulties. Firstly, there is no explicit mathematical relationship linking the decision variables (genetic parameters and management practices) to the models' outputs (criteria and constraints). Secondly, these relationships are probably nonlinear as suggested by the results of simulation techniques. Thirdly, it is impossible to test all the candidate combinations due to their huge number. Fourthly, the fitness landscapes to be explored are usually complex and have many local optima (Messina et al. 2009). Therefore, model-based ideotype-design triggers a very difficult optimisation problem. In order to proceed, we need efficient optimisation algorithms able to tackle these difficulties that resist classical optimisation and simulation methods. In a large number of domains, nature-inspired optimisation algorithms (e.g., genetic algorithms, ant colonies, particle swarm optimisation algorithms) appear efficient to solve difficult optimisation problems. These algorithms allow exploring highly dimensional solution spaces in a reasonable computation time. In addition, they do not require any derivative information, and can address the complex multi-objective optimisation problems (e.g., very large search spaces, uncertainty, noise, disjoint Pareto curves). As a result, the use of nature-inspired algorithms has recently expanded in the domain of model-based ideotype design (He et al. 2012; Kadrani et al. 2012; Letort et al. 2008; Qi et al. 2010; Quilot-Turion et al. 2012) and optimisation of management scenarios (Grechi et al. 2012). These methods provide a set of diversified solutions to decision-makers and let them have the final choice of the best-suited trade-off between criteria.

We will illustrate now the use of nature-inspired multi-objective optimisation algorithms to design ideotypes. Our illustration concerns the peach fruit susceptibility to a pathogenic fungus (*Monilinia* spp.), responsible of the peach brown rot. This infection is largely occurring through fruit-wounds and it is proven that cuticular cracks play a major role for fungal infection (Gibert et al. 2009). Cuticular crack density has been shown to rise with fruit growth rate, which in its turn varies with management practices such as irrigation regimes and fruit thinning (Gibert et al. 2010). The cracks represent large opportunities for fungal infection and contribute also to fruit water losses by transpiration, influencing thereby fruit growth and quality build-up. *Monilinia* spp. causes in France a rising problem mainly due to the current reduction of fungicide usage in orchards for ecological, economic and human-health reasons. This short description illustrates clearly the conflicting objectives of the peach growers since the production of big fruits, which are generally the sweetest in taste, leads inevitably to elevated *Monilinia* attacks. In this case, growers should be interested to find an adequate trade-off between quality and fruit health performances. It is needless to say that this trade-off depends also on the

environmental conditions. Thus, overall we must search for the best possible exploitation of G×E×M interactions, seeking specific adaptation of some genotypes to target population of environments (TPE) (Cooper and Byth 1996). The idea is to cluster crop environments according to their limiting factors and the frequency of their occurrence. Within a class of TPE, a given genotype may have an expected and stable behaviour. Thus, the optimisation step must be performed in various TPE differing by cultural practices and/or climate profiles so that different ideotypes can emerge. Therefore, the definition of relevant target environments in which the limiting factors are well identified, represents a major step to focus the subsequent ideotype design.

As an attempt to sort out this problem, we used the "Virtual Fruit" model to design peach ideotypes with large fruit mass, good sweetness, and low density of skin cracks, i.e., low sensitivity to brown rot.

As mentioned before, the SA performed on the model identified the main parameters affecting some important outputs of the Virtual Fruit model, i.e., fruit mass, sweetness, crack density. From this SA, we selected six parameters affecting the processes of vegetative activity, i.e., fruit growth, sugar metabolism, stone-pulp partition, transpiration and water fluxes. These parameters are considered to be genotype-dependent and mostly independent of the environment. Consequently, each genotype may be considered hereafter as a set of six genetic parameters. We ran the model under two contrasted modalities of irrigation regimes and thinning intensities in order to analyse the impact of cultural practices on the optimised solutions. Irrigation and thinning practices are particularly relevant in the peach system due to their ecological impact and labour requirement. Moreover, both have large effects on fruit yield, quality and cuticular cracking. The optimised solutions emerged from the coupling of the model with different nature-inspired multi-objective optimisation algorithms (for details, see Kadrani et al. 2012, 2013; Ould-Sidi et al. 2012; Quilot-Turion et al. 2012). All algorithms provided a large diversity of solutions, among which it was possible to choose the best-suited trade-off between criteria according to a particular objective. Our results confirmed that the solution domain was strongly constrained by the fruit mass and the microcrack characteristics and that there was a strong antagonism between the criteria considered (Fig. 4.6). For example, we identified solutions matching current breeding schemes where fruit mass is the sole rated criterion. However, we also identified some interesting optimised solutions representing a breakthrough in the current schemes, which could be considered in the future as they do favour organoleptic quality and environment-friendly practices. The global impact of crop load appeared high on fruit mass and crack density and moderate on sweetness. Irrigation regime displayed a small impact on the three criteria in case of low crop load but gave rise to contrasting phenotypes in case of high crop load. We found that whatever the scenario, ideotypes with lower fruit mass exhibited greater sweetness and had lower crack density on their skin. According to the 'Virtual Fruit' model, we may predict that it is not possible to get a set of six genetic parameters leading to ideotypes of very high fruit mass and sweetness but low sensitivity to brown rot. The figure reveals that commercial standards picking out cultivars with big fruits promote high

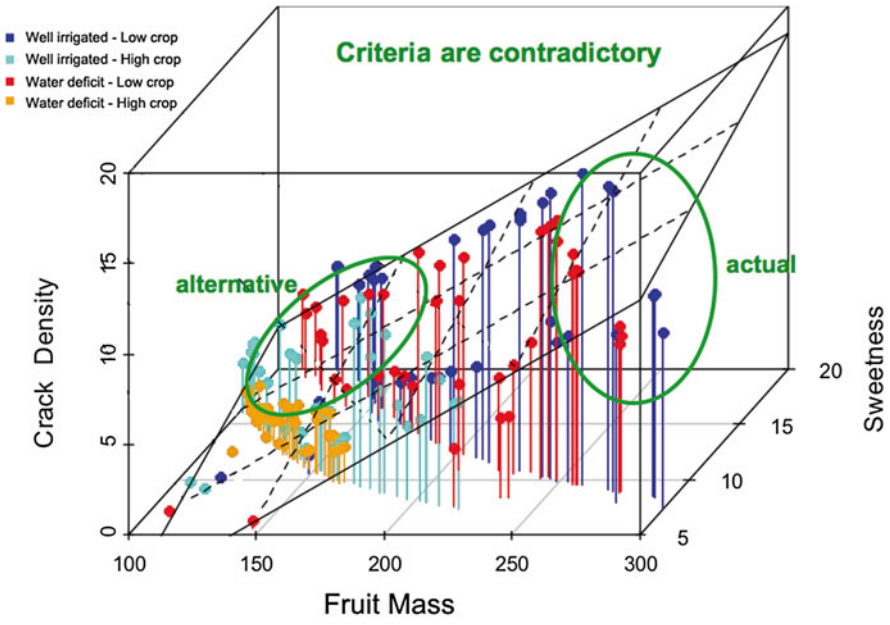


Fig. 4.6 Virtual Fruit outputs for the set of final solutions for four environmental scenarios (two levels of crop load and two water regimes). The optimization was done using the NSGA-II (Non-dominated Sorting Genetic Algorithm II) algorithm, which explores complex search spaces in a reasonable computation time. More details can be found in Quilot-Turion et al. (2012). Cuticular crack density is plotted against fruit mass and sweetness. The highlighted solutions (green ovals) identify the actual breeding strategies versus putative alternative strategies devising a compromise between the three targeted traits

sensitivity to brown rot and low fruit sweetness. According to our simulations an interesting compromise would be to breed for cultivars with lower fruit mass, obtained for example by decreasing the hydraulic conductance of tissues (one of the six parameters tested) in order to get sweet fruits with low sensitivity to brown rot. This would be an acceptable trade-off favouring organoleptic quality and environmentally friendly practices.

The results of these pioneer studies illustrate the value of the multi-objective optimisation approach. The next important step will be applying this approach to a prospective search of ideotypes adapted to future climates or to seek opportunities for crop expansions into new areas. This is particularly relevant in the case of perennial crops that are settled in an orchard for decades. Historical data or outputs from climate models could be used as inputs in the Virtual Fruit model. The output simulations could thus help identify the main traits of genotype adaptation to changing climates for the future.

However, an important initial step is needed before using this approach to such a prospect. Indeed, a main weakness of the methodology is the current lack of quantitative relationships between genes and model parameters. In fact, this approach

simply provides a picture of the optimised space of solutions from the viewpoint of the system functioning under the constraint of biophysical limiting factors. Presently the suggested solutions represent ideal genotypes that breeders may not be able to create. In order to produce more realistic genotypes, genetic constraints (such as pleiotropic and epistatic effects, and G×E interactions; see Chap. 1 of this book by Baldazzi et al.) must be integrated into the optimisation scheme.

4.5 Conclusion

To breed for cultivars that can adapt to changing climates to sustain or to increase yield and product quality, we need a better understanding of the complex interactions among the numerous factors that drive resource acquisition, distribution and storage in response to environmental stimuli. Moreover, there is an urgent need to identify and track key genes involved in plant adaption to stress-prone environments. The PBSMs described in this chapter have proven very efficient for unravelling plant complexity and plasticity in response to environmental stimuli or genetic perturbations. In addition, we have illustrated how PBSMs can be used as powerful tools for phenotyping plants at the process level in a comprehensive and "costless" way. As a consequence, these models should also be regarded as high-throughput phenotyping platforms that may complement the more expensive genetic, proteomic or metabolomic platforms. Indeed our assessment of the Virtual Fruit model and the root system architecture model shows clearly that PBSMs may help us to disentangle the links between genotypes (i.e. a set of parameters) and phenotypes. On the one hand, we can identify optimized sets of genetic parameters to achieve a target phenotype. On the other hand, models help us understand how a given phenotype may emerge from a specific genotype. Over the last decade, near-isogenic lines, mutants, transgenic and mapping populations have been developed for several traits related to fruits quality. Such materials give scientists the opportunity to evaluate the hypotheses introduced in the models and to highlight important regulation loops.

Nevertheless, we are still far from the rise of new cultivars based on model-designed ideotypes. One main reason is the gap between model parameters and genes or physiological functions. To progress in this field, more work is expected in order to fill the gap between genetic information concerning the traits and processes included in models. Moreover, PBSMs must be refined by more mechanistic details able to enlarge their ability to simulate the complexity of plant and organ functioning. Adding genetic and genomic information on gene actions and interactions into PBSMs will help modellers to unravel and strengthen physiological assumptions and equations in their models. It will also help to reduce some uncertainty about the genetic knowledge caused by environment interactions. In our opinion, it is not essential to understand fully a trait genetics and physiology to model it at an operational level. Nevertheless, modellers are seeking adequate data sets desperately, offering time series and accurate description and characterisation of cultural conditions and genotypes, so as to parameterise and evaluate their PBSMs. In this

prospect, it would be advisable to establish a large database including phenotype data measured at plant, organ and process levels and cross this information with molecular and genetic databases.

References

Asseng S, Turner NC, Ray JD, Keating BA (2002) A simulation analysis that predicts the influence of physiological traits on the potential yield of wheat. Eur J Agron 17:123–141

Baldazzi V, Bertin N, de Jong H, Génard M (2012) Towards multiscale plant models: integrating cellular networks. Trends Plant Sci 17:728–736

Blower SM, Dowlatabadi H (1994) Sensitivity and uncertainty analysis of complex-models of disease transmission – an HIV model, as an example. Int Stat Rev 62:229–243

Boote KJ, Jones JW, Batchelor WD, Nafziger ED, Myers O (2003) Genetic coefficients in the CROPGRO-soybean model: links to field performance and genomics. Agron J 95:32–51

Bylesjö M, Eriksson D, Kusano M, Moritz T, Trygg J (2007) Data integration in plant biology: the O2PLS method for combined modeling of transcript and metabolite data. Plant J 52:1181–1191

Chapman S, Cooper M, Podlich D, Hammer G (2003) Evaluating plant breeding strategies by simulating gene action and dryland environment effects. Agron J 95:99–113

Cooper M, Byth DE (1996) Understanding plant adaptation to achieve systematic applied crop improvement – a fundamental challenge. In: Cooper M, Hammer GL (eds) Plant adaptation and crop improvement. CAB International, Wallingford, pp 5–23

de Dorlodot S, Forster B, Pagès L, Price A, Tuberosa R, Draye X (2007) Root system architecture: opportunities and constraints for genetic improvement of crops. Trends Plant Sci 12:474–481

Dunbabin VM, Postma JA, Schnepf A, Pagès L, Javaux M, Wu LH, Leitner D, Chen YL, Rengel Z, Diggle AJ (2013) Modelling root-soil interactions using three-dimensional models of root growth, architecture and function. Plant Soil 372:93–124

Fischer KS (1996) Improving cereals for the variable rainfed system: from understanding to manipulation. In: Singh VP, Singh RK, Singh BB, Zeigler RS (eds) Proceedings of the international conference on stress physiology of rice. NDUAT/International Rice Research Institute, Los Baños, pp 1–10. 28 Feb–5 Mar 1994, Lucknow

Génard M, Gouble B (2005) ETHY. A theory of fruit climacteric ethylene emission. Plant Physiol 139:531–545

Génard M, Bertin N, Borel C, Bussières P, Gautier H, Habib R, Lechaudel M, Lecomte A, Lescourret F, Lobit P, Quilot B (2007) Towards a virtual fruit focusing on quality: modelling features and potential uses. J Exp Bot 58:917–928

Génard M, Bertin N, Gautier H, Lescourret F, Quilot B (2010) Virtual profiling: a new way to analyse phenotypes. Plant J 62:344–355

Gibert C, Lescourret F, Génard M, Vercambre G, Pastor AP (2005) Modelling the effect of fruit growth on surface conductance to water vapour diffusion. Ann Bot Lond 95:673–683

Gibert C, Chadoeuf J, Nicot P, Vercambre G, Génard M, Lescourret F (2009) Modelling the effect of cuticular crack surface area and inoculum density on the probability of nectarine fruit infection by *Monilinia laxa*. Plant Pathol 58:1021–1031

Gibert C, Génard M, Vercambre G, Lescourret F (2010) Quantification and modelling of the stomatal, cuticular and crack components of peach fruit surface conductance. Funct Plant Biol 37:264–274

Grechi I, Ould-Sidi MM, Hilgert N, Senoussi R, Sauphanor B, Lescourret F (2012) Designing integrated management scenarios using simulation-based and multi-objective optimization: application to the peach tree-*Myzus persicae* aphid system. Ecol Model 246:47–59

Habekotté B (1997) Options for increasing seed yield of winter oilseed rape (*Brassica napus* L.): a simulation study. Field Crop Res 54:109–126

Hammer GL, Kropff MJ, Sinclair TR, Porter JR (2002) Future contributions of crop modelling – from heuristics and supporting decision making to understanding genetic regulation and aiding crop improvement. Eur J Agron 18:15–31

Hammer GL, Chapman S, van Oosterom E, Podlich DW (2005) Trait physiology and crop modelling as a framework to link phenotypic complexity to underlying genetic systems. Aust J Agr Res 56:947–960

Haverkort AJ, Grashoff C (2004) IDEOTYPING-POTATO a modelling approach to genotype performance. In: MacKerron DKL, Haverkort AJ (eds) Decision support systems in potato production. Wageningen Academeic, Wageningen, pp 198–211

He JQ, Le Gouis J, Stratonovitch P, Allard V, Gaju O, Heumez E, Orford S, Griffiths S, Snape JW, Foulkes MJ, Semenov MA, Martre P (2012) Simulation of environmental and genotypic variations of final leaf number and anthesis date for wheat. Eur J Agron 42:22–33

Hennig L (2007) Patterns of beauty – omics meets plant development. Trends Plant Sci 12:287–293

Herndl M, C-g S, Wang P, Graeff S, Claupein W (2007) A model based ideotyping approach for wheat under different environmental conditions in North China plain. Agric Sci China 6:1426–1436

Jacques J (2011) Pratique de l'analyse de sensibilité: comment évaluer l'impact des entrées aléatoires sur la sortie d'un modèle mathématique. Université de Lille, s.n.:1–14

Kadrani A, Ould-Sidi M-M, Quilot-Turion B, Génard M, Lescourret F (2012) Particle swarm optimization to design ideotypes for sustainable fruit production Systems. Int J Swarm Intell Res (IJSIR) 3:1–19

Kadrani A, Ould-Sidi MM, Quilot-Turion B, Génard M, Lescourret F (2013) Comparison of evolutionary and swarm intelligence-based approaches in the improvement of fruit quality. International symposium on operational research and applications (ISORAP2013), Marrakesh, 8–10 May 2013

Kopka J, Fernie A, Weckwerth W, Gibon Y, Stitt M (2004) Metabolite profiling in plant biology: platforms and destinations. Genome Biol 5:109

Lescourret F, Génard M (2005) A virtual peach fruit model simulating changes in fruit quality during the final stage of fruit growth. Tree Physiol 25:1303–1315

Letort V, Mahe P, Cournede PH, De Reffye P, Courtois B (2008) Quantitative genetics and functional-structural plant growth models: Simulation of quantitative trait loci detection for model parameters and application to potential yield optimization. Ann Bot Lond 101:1243–1254

Lobit P, Génard M, Wu BH, Soing P, Habib R (2003) Modelling citrate metabolism in fruits: responses to growth and temperature. J Exp Bot 54:2489–2501

Messina CD, Hammer GL, Dong Z, Podlich D, Cooper M (2009) Modelling crop improvement in a GxExM framework via gene-trail-phenotype relationships. In: Sadras VO, Calderini D (eds) Crop physiology: applications for genetic improvement and agronomy. Elsevier, The Netherlands, pp 235–265

Morris MD (1991) Factorial sampling plans for preliminary computational experiments. Technometrics 33:161–174

Ould-Sidi MM, Kadrani A, Quilot-Turion B, Lescourret F and Génard M (2012) Compromising NSGA-II performances and stopping criteria: case of virtual peach design. In: 4th international conference on metaheuristics and nature inspired computing, META'2012, Port Kintaoui

Pagès L (2011) Links between root developmental traits and foraging performance. Plant Cell Environ 34:1749–1760

Pagès L, Ariès F (1988) SARAH – a simulation-model for growth, development and architecture of root systems. Agronomie 8:889–896

Pagès L, Vercambre G, Drouet JL, Lecompte F, Collet C, Le Bot J (2004) Root typ: a generic model to depict and analyse the root system architecture. Plant Soil 258:103–119

Pagès L, Bruchou C, Garre S (2012) Links between root length density profiles and models of the root system architecture. Vadose Zone J 11:4
Pagès L, Bécel C, Boukcim H, Moreau D, Nguyen C, Voisin AS (2014) Calibration and evaluation of ArchiSimple, a simple model of the root system architecture. Ecol Model 290:76–84
Peck SL (2004) Simulation as experiment: a philosophical reassessment for biological modeling. Trends Ecol Evol 19:530–534
Qi R, Ma YT, Hu BG, de Reffye P, Cournede PH (2010) Optimization of source-sink dynamics in plant growth for ideotype breeding: a case study on maize. Comput Electron Agric 71:96–105
Quilot B, Kervella J, Génard M, Lescourret F (2005) Analysing the genetic control of peach fruit quality through an ecophysiological model combined with a QTL approach. J Exp Bot 56:3083–3092
Quilot-Turion B, Ould-Sidi M-M, Kadrani A, Hilgert N, Génard M, Lescourret F (2012) Optimization of parameters of the 'Virtual Fruit' model to design peach genotype for sustainable production systems. Eur J Agron 42:34–48
Reymond M, Muller B, Leonardi A, Charcosset A, Tardieu F (2003) Combining quantitative trait loci analysis and an ecophysiological model to analyze the genetic variability of the responses of maize leaf growth to temperature and water deficit. Plant Physiol 131:664–675
Saltelli A, Tarantola S, Chan KPS (1999) A quantitative model-independent method for global sensitivity analysis of model output. Technometrics 41:39–56
Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, Saisana M, Tarantola S (2008) Global sensitivity analysis. The primer. Willey, Chichester
Stewart DW, Cober ER, Bernard RL (2003) Modeling genetic effects on the photothermal response of soybean phenological development. Agron J 95:65–70
Struik PC, Yin X, de Visser P (2005) Complex quality traits: now time to model. Trends Plant Sci 10:513–516
Tardieu F (2003) Virtual plants: modelling as a tool for the genomics of tolerance to water deficit. Trends Plant Sci 8:9–14
Tomita M, Hashimoto K, Takahashi K, Shimizu TS, Matsuzaki Y, Miyoshi F, Saito K, Tanida S, Yugi K, Venter JC, Hutchison CA (1999) E-CELL: software environment for whole-cell simulation. Bioinformatics 15:72–84
Trewavas A (2003) Aspects of plant intelligence. Ann Bot Lond 92:1–20
Trewavas A (2004) Aspects of plant intelligence: an answer to Firn. Ann Bot Lond 93:353–357
Trewavas A (2006) A brief history of systems biology – "Every object that biology studies is a system of systems." Francois Jacob (1974). Plant Cell 18:2420–2430
Uys L, Botha FC, Hofmeyr JHS, Rohwer JM (2007) Kinetic model of sucrose accumulation in maturing sugarcane culm tissue. Phytochemistry 68:2375–2392
Welch SM, Dong ZS, Roe JL, Das S (2005) Flowering time control: gene network modelling and the link to quantitative genetics. Aust J Agr Res 56:919–936
Wu BH, Génard M, Lobit P, Longuenesse JJ, Lescourret F, Habib R, Li SH (2007) Analysis of citrate accumulation during peach fruit development via a model approach. J Exp Bot 58:2583–2594
Yin X, Chasalow SD, Dourleijn CJ, Stam P, Kropff MJ (2000) Coupling estimated effects of QTLs for physiological traits to a crop growth model: predicting yield variation among recombinant inbred lines in barley. Heredity 85:539–549
Yin X, Struik PC, Tang JJ, Qi CH, Liu TJ (2005) Model analysis of flowering phenology in recombinant inbred lines of barley. J Exp Bot 56:959–965

Chapter 5
Heuristic Exploration of Theoretical Margins for Improving Adaptation of Rice through Crop-Model Assisted Phenotyping

Delphine Luquet, Camila Rebolledo, Lauriane Rouan, Jean-Christophe Soulie, and Michael Dingkuhn

Abstract Crop modeling in support of breeders' decisions on selection criteria can benefit from the new global focus on phenomics because it provides new information on existing genetic diversity for useful traits. This study attempted an *in silico* prediction of margins for genetic improvements of early vigor (biomass produced during vegetative growth) and drought resistance combined, based on virtual recombination of several traits (here *syn.* model parameters) within ranges of trait variation observed in a panel of diverse rice genotypes. The Ecomeristem model was parameterized by multi-parameter optimization procedures applied to observed datasets for 136 rice genotypes. The traits within the observed ranges were then recombined *in silico* to generate a virtual population of 9000 individuals. Simulations of real and virtual phenotypes under three water treatments, using finite water resources during stress cycles, indicated strong and similar trade-offs between constitutive vigor and drought resistance in both real and virtual, recombinant populations. A substantial margin for potential genetic improvement of vigor with unchanged drought resistance was predicted, drawing chiefly from structural growth and development traits that would increase internal demand for assimilates (larger and thicker leaves, increased leaf appearance rates). Increased vigor would not necessarily require greater photosynthetic potential per se. However, improved drought resistance with unchanged constitutive vigor would require greater water economy (increased photosynthetic potential and limited water use, therefore higher transpiration efficiency) and greater tolerance of leaf extension and gas exchange

D. Luquet (✉) • L. Rouan • J.-C. Soulie
UMR AGAP, CIRAD, Montpellier, France
e-mail: luquet@cirad.fr

C. Rebolledo
CIAT, Agrobiodiversity, AA 6713 Cali, Colombia

M. Dingkuhn
UMR AGAP, CIRAD, Montpellier, France

CESD Division, IRRI, DAPO Box 7777, Metro Manila, Philippines

X. Yin, P.C. Struik (eds.), *Crop Systems Biology*,
DOI 10.1007/978-3-319-20562-5_5

rates to drought, while tillering ability should be limited in favor of larger and thicker leaves. These results carry significant uncertainty because they predict virtual genotypes and their phenotypes, based on simple assumptions in the model (namely on gas exchange) and in genetics (free, additive trait combinability). But the approach is innovative and may eventually help developing ideotypes drawing from information of existing diversity and integrative modeling of phenotypes.

5.1 Introduction

5.1.1 Integrative Modeling of the Genetic and Environmental Control of Complex Traits

The challenge of using integrative modeling to unravel the genetic and environmental control of complex traits, such as crop biomass production, is debated since about two decades (Dingkuhn et al. 2005; Hammer et al. 2004, 2005; Yin et al. 2004). These authors proposed that whole plant functioning is the pivotal biological scale to understand the effect of key genes, gene networks and related biological processes on the diversity and plasticity of phenotypes and thus of crop performance. Its analysis through integrative approaches is needed to be useful for prebreeding research and breeding, as crop ideotypes are also described at the whole-plant and plant population scales (Araus et al. 2002; Chapman et al. 2003).

Systems biology can integrate information and knowledge at organizational scales from molecular to plant or even crop level (Baldazzi et al. 2013; Keurentjes et al. 2011; Yin and Struik 2010). The complexity of plant systems biology requires integrative modeling that situates the component processes in a dynamic web of feedbacks and trade-offs, involving many interacting entities (Baldazzi et al. 2013; also see Chap. 1 of this book by Baldazzi et al.). Efforts made in this sense explored the genetic and environmental control of phenotypes using heuristics (Dingkuhn et al. 2005; Hammer et al. 2002). They require high-throughput methodologies both for the physical phenotyping process and the subsequent model application (Luquet et al. 2012b; Quilot-Turion et al. 2012). However, while heuristics using integrative modeling has been discussed extensively and modeling approaches have specifically been developed for complex systems (Cooper et al. 2002, 2014; Quesnel et al. 2009), their practical use for genetic analysis of complex traits is not trivial (Dingkuhn et al. 2005; Hammer et al. 2010; Xu et al. 2011). Pioneering work using fitted model parameters as phenotypic traits for genetic QTL (Quantitative Trait Locus) studies were so far mainly reported for single-process models or models focused on a specific plant organ, such as a leaf expansion model for maize (Reymond et al. 2004), peach fruit quality (Quilot et al. 2005), or specific leaf area (SLA) of barley (Yin et al. 1999). In the case of the maize leaf expansion model, the phenomics and heuristics were conducted at the single-process and single-organ scale, and only *a posteriori* scaled up to the crop level (Chenu et al. 2009). A variant of this approach was performed on rice (Gu et al. 2014) and barley

(Yin et al. 2000), whereby certain crop traits were phenotyped for a mapping population and the QTL effects directly used to substitute the corresponding crop parameter effects in the complex crop model. In both approaches, the complex crop model was used as a tool to predict phenotypes of recombinant genotypes, but it was not used as a phenotyping tool per se.

In theory, the same complex crop model can be used to assist the phenotyping (by fitting parameters for traits that are difficult to measure directly), for QTL analyses on the basis of these parameters, and for predicting recombinant phenotypes. The feasibility and added value of this approach remains to be demonstrated.

5.1.2 Mechanisms of Early Vigor and Their Interaction with Drought

When grown under rainfed conditions, the rice crop intermittently encounters drought periods, with potentially severe impact on grain yield (Babu et al. 2001; Jongdee et al. 2006; Kamoshita et al. 2004). The variability of tropical rainy seasons is expected to be aggravated by global warming. Drought events during reproductive development directly affect fertile panicle and spikelet number and grain filling, but early drought events are also of concern (Kamoshita et al. 2004). At early stages, the rice seedlings are fragile and drought may impede the production of leaf area and structural biomass needed to support sufficient tiller, panicle and grain numbers (Richards and Lukacs 2001). Early biomass and leaf area generation, called early vigor, is also needed to colonize space, acquire resources, and compete with weeds (Namuco et al. 2009).

For these reasons, early vigor is important both from an ecological (survival) and agronomic (performance) point of view. But early vigor also accelerates resource use including the soil water reserve and thereby increases the probability of drought occurrence during later growth periods. Early vigor and water use are thus intrinsically linked and cause trade-offs with drought avoidance, but the extent of these trade-offs may depend on the environment, cultural practices and the specific trait combinations expressed in a genotype. In the present study, we are particularly interested in the latter.

Both early vigor and drought resistance (composed of tolerance and avoidance traits) are complex traits. Vigor in the absence of drought involves not only processes of light capture and carbon (C) acquisition (sources), but also developmental processes of organ deployment (constituting sinks), and thus involves source-sink regulation. To what extent a vegetative plant can be source or sink limited is an ongoing debate (Dingkuhn et al. 2007), particularly in the presence of drought (Pallas et al. 2013; Pantin et al. 2011). Sink limitation can be of ecological advantage, by conserving a finite external resource, storing internal reserves and limiting plant exposure to stress (but rendering the crop more vulnerable to weed competition). Source limited growth in the absence of stress may be considered a more aggressive strategy but increases exposure if stress occurs. Source limited growth under drought can be directly caused by stress effects on gas exchange or secondarily by leaf

senescence. Organogenetic (developmental) sinks tend to be down-regulated under stress and their survival, in terms of meristem integrity, is vital (Tisne et al. 2010). Therefore, sink maintenance under stress may sometimes be of agronomic advantage but is risky in drought prone ecologies.

It was suggested that up to a moderate drought level plants are frequently sink limited, showing a positive nonstructural carbohydrate (NSC) balance (Luquet et al. 2008; Pantin et al. 2011), but under more severe stress, plants become C source limited. Accordingly, growth maintenance under drought is a subtle trade-off between C source and sink regulation and depends both on the drought type and the species (Pallas et al. 2013).

In rice, early vigor and its interaction with drought show substantial genetic diversity (Rebolledo et al. 2012a, b, 2013). A phenotypic study of a large panel of diverse accessions revealed a negative correlation between early vigor (both in terms of biomass accumulation and organogenetic development rate) and leaf starch content. Plants having low vigor and high starch concentrations were interpreted as being sink limited. High vigor, where it was observed, could be related to organ number (deployment rate of leaves and tillers) or organ size, whereby organ size and number were negatively correlated across genotypes. In general, early vigor was negatively correlated with drought tolerance, which was defined as the capacity of seedlings to maintain biomass accumulation under a given soil water deficit. Therefore, the negative trade-off between constitutive vigor and drought tolerance was not a result of earlier water reserve depletion by the vigorous types, but of other processes. It thus appears that selecting for vigor may incur loss of drought tolerance, and vice versa.

5.1.3 Modeling Early Vigor and Drought Interactions

While at the organ level the switch from a metabolic to hydraulic control of growth during drought development was experimentally demonstrated (Pantin et al. 2011), an integrative analysis of such processes at the whole-plant scale remains difficult (Luquet et al. 2008). Regulation of C source-sink and water relationships varies with genotype. Several modeling schools emerged to address this question. Tardieu et al. (2011) described two approaches commonly used in plant growth models: (i) plant growth under water deficit is driven by integrative plant variables such as plant carbon status (Yan et al. 2004), or (ii) parallel mechanisms affect plant expansive growth (hormonal and hydraulic signals) and biomass accumulation (stomatal conductance, photosynthesis). The latter approach was implemented in many crop models (Brisson et al. 2003; Hammer et al. 2010). A novel generation of models (Pallas et al. 2013) considers combined effects of water status and sink/source relations on organ deployment and expansion. For grasses, this concept is implemented in Ecomeristem (Luquet et al. 2006), a model that simulates plant morphogenesis and phenotypic plasticity as modulated by assimilate status (as a whole-plant ratio of aggregate supply to aggregate demand) and water status (as a function of fraction of transpirable soil water in the root zone, FTSW). Water status thereby affects

both assimilation rate (affecting source) and organ expansion (affecting sink), and the resulting source/sink ratio feeds back on deployment of new organs and their potential size, based on a dynamic representation of plant topology at organ level. The originality of the model resides in the assumption of a single pool of C assimilate that many entities (organs) compete for, the resulting level of competition then feeding back on new growth commitments in terms of organ deployment and potential size. Transient source/sink imbalances are buffered by a reserve pool within a day, and compensated by organ initiation or senescence. Recently, Ecomeristem was used to phenotype a genetically diverse rice panel (136 accessions) for early vigor under abundant water supply (Luquet et al. 2012b) and under drought (Luquet et al. 2012a). Parameter multi-fitting was performed for several environments by optimization procedures using a genetic algorithm. The model was able to reproduce and predict key trait combinations observed within the population, such as a close association of early vigor with organogenetic development rate (inverse of phyllochron), a negative association of vigor with the size of transient NSC pools, and strong trade-offs between constitutive vigor and drought tolerance (Rebolledo et al. 2012b, 2013).

Genotypic model parameter values can be regarded as traits and used to study their genetic control by way of genome-wide association studies *(GWAS)*. This work is not presented here. However, the calibrated model can also be used to predict phenotypes in response to variation in environment. This approach can be taken one step further by simulating virtual genotypes that differ from the observed ones in the values or combinations of values for specific traits (Chenu et al. 2009; Yin et al. 1999). Ultimately, the approach potentially enables predicting improved ideotypes, and if strong associations of model parameters with genomic regions are found via GWAS, molecular markers can be developed for the marker-assisted breeding for such ideotypes.

The present study aimed to (i) simulate with Ecomeristem the performance of a rice population of 136 genotypes (for which the model was previously trained, Luquet et al. 2012a, b) under three environmental conditions (well-watered and two drought levels); and (ii) compare it with the performance of a virtual population of "recombinant" genotypes, generated by randomly combining parameter values within the ranges met in the original population. On this basis, we try to explore potential margins for improving early vigor, improving drought tolerance and limiting trade-offs between vigor and tolerance.

5.2 Materials and Methods

5.2.1 Ecomeristem Model

The Ecomeristem model simulates plant vegetative morphogenesis of rice, sorghum and sugarcane. The model was described in previous studies on rice plants under non-stress (Luquet et al. 2006) and drought conditions (Luquet et al. 2012a; Pallas et al. 2013). Regarding the rice model for the vegetative phase (before internode

elongation), phytomer initiation rate is scheduled by a potential plastochron (*Plasto*, genotypic parameter). It is equal to the phyllochron, i.e., the duration (in °Cd) of the expansion phase of a leaf from its tip appearance until the next leaf-tip appearance (a relationship specific to rice) (Nemoto et al. 1995). Once initiated, an organ *n* is pre-dimensioned. Its potential final size is computed as the final length of leaf (*n*-1) incremented by a genotypic parameter (*MGR*: Meristem Growth Rate). *MGR* quantifies the ability of the vegetative meristem to produce successive leaves with an increasing size. An allometric coefficient is used to translate leaf length into width (Luquet et al. 2006). Once pre-dimensioned, a leaf expands at a Leaf Expansion Rate (LER, $cm^2.°Cd^{-1}$) equal to the ratio between potential final leaf length and expansion duration.

The root compartment is only simulated in terms of biomass, i.e., as a bulk compartment with a daily growth demand computed proportionally to that of the shoot and depending on plant phenology. The sum of daily potential leaf and root growth constitutes plant demand for carbohydrates (C). C supply is computed daily at crop level (scaled to plant level by dividing by planting density) using the commonly used big-leaf approach based on a light extinction rate (*Kdf*) and a light conversion efficiency (*epsib*, g MJ^{-1}). The daily ratio between plant C supply and demand constitutes a state variable named Ic, an index of plant internal competition for C, which modulates many processes simulated by the model. Both leaf dimensioning (at initiation time) and LER are down-regulated if Ic is inferior to 1. Tillering is also regulated by Ic, depending on a genotypic threshold parameter *Ict* (Ic threshold above which tillering occurs). A genotype with low *Ict* tillers more easily than one with high *Ict*. A particularity of Ecomeristem is also the modeling of a whole-plant C storage compartment governed by Ic: C reserves are fed into it if Ic>1 and mobilized if Ic<1. In the latter case, if C supply falls short of demand even after reserve mobilization, the senescence of the oldest leaf on the plant is accelerated.

Areal leaf expansion is translated into structural dry weight demand using a leaf rank dependent value of SLA (specific leaf area, surface area/structural dw) computed by a logarithmic equation attenuated by a slope parameter, *SLAp*. As mentioned above, Ecomeristem uses the state variable FTSW do drive plant functioning under drought (Luquet et al. 2012a). FTSW impacts directly on leaf expansion and transpiration rates according to a bilinear broken-stick equation dependent on threshold parameters (*ThresLER* and *ThresTransp*). Transpiration reduction rate impacts on potential C assimilation through Cstr, ratio between actual and potential plant leaf transpiration, according to a power parameter (*epsibsens*, cf. equation in Table 5.1).

5.2.2 Model Calibration on Rice Diversity Panel

A rice diversity panel of 136 individuals was studied in a greenhouse experiment regarding traits related to early vigor and its drought regulation (IRRI, Philippines 2013). This panel contained mainly tropical japonica but also other groups (19 indica, 5 temperate japonica, 2 aromatic and 2 Aus). Early vigor was considered as the capacity to accumulate shoot biomass rapidly. Drought tolerance was

Table 5.1 Definition, unit and ranges of values of Ecomeristem model parameters: bold values: ranges used for model calibration for each of the three replicates of each of the 136 accessions present in the rice diversity panel; other range: interval of values eventually explored by the diversity panel (considering average of three replicates per genotype). P values indicate the significance for a replicate effect on parameter value

Parameters	Identification	Range used for calibration vs. range found in the rice panel
epsib	Light conversion efficiency ($g.MJ^{-1}$)	**3–6** vs. 3.53–5.92 ($p<0.05$)
Ict	Ic threshold enabling tillering	**0.5–3** vs. 0.65–1.99
Kcpot	Potential leaf transpiration	**7–14** vs. 7–13.62 ($p<0.05$)
MGR	Meristem growth rate (cm)	**3–5** vs. 3.76–5
Plasto	Plastochron (°Cd)	**40–80** vs. 42.56–68.99
epsibsens	Power parameter to reduce epsib proportionally to water stress level as such: $epsib*Cstr^{epsibsens}$ (where Cstr is defined by Cstr = actual/potential leaf transpiration rate)	**0.5–1.5** vs. 0.5–1.5
SLAp	Parameter controling specific leaf area variation with leaf rank	**20–80** vs. 25.38–80
thresLER	Threshold FTSW for leaf expansion rate reduction	**0–1** vs. 0.14–1
thresTransp	Threshold FTSW for leaf transpiration rate reduction	**0.4–1** vs. 0.4–0.89 ($p<0.05$)

considered as the maintenance of green-shoot dry weight (GSDW) accumulation under drought. For the latter purpose, a drought treatment was compared with a control (well-watered) one. Each genotype in each treatment was represented by three potted plants (replicates). This experiment was extensively described by Rebolledo et al. (2013). The drought treatment was created by stopping irrigation for each accession individually at the appearance of leaf 6 on the main stem. Pots (1-liter pots, size chosen to avoid any genotypic difference in terms of soil exploration by roots) were watered at field capacity till that time and then kept dry-down until a severe level of stress was reached (FTSW of 0.2). Water stress was quantified by the Fraction of Transpirable Soil Water (FTSW) that was measured gravimetrically, as well as plant leaf transpiration rate. Measured traits were in both treatments, the last expanded leaf size and rank on the main stem, tillering, leaf senescence rate, at least three times during the experiment. These phenotypic data, together with corresponding meteorological data (daily photosynthetically active radiation, air temperature at the base of plants and evaporative demand) were used to calibrate Ecomeristem for each accession in the studied panel. Parameter estimation was performed using the genetic algorithm method Rgenoud (Sekhon and Mebane 2011). For each accession and each replicate (3) in the panel, 9 genotypic parameters were estimated: 6 morphogenetic and 3 drought response parameters (see Table 5.1). These parameters were addressed because they control traits known to vary across genetic materials in the panel (Rebolledo et al. 2012a, b, 2013) and because they

were confirmed to impact simulated shoot biomass accumulation (evaluated through model sensitivity analysis in Luquet et al. 2012a). Other parameters were fixed at a value specific to rice already used in previous studies (Luquet et al. 2012a). Data used for fitting the model and estimating parameter values were those measured at stress onset and at the end of the stress period: GSDW, tiller number, total leaf number on the plant, green and total leaf number on the main stem, length and width of the last ligulated leaf on the main stem, and FTSW. Parameter values obtained for each replicate of a given genotype were then averaged, which was made possible due to the absent or small replicate effects observed (cf. Table 5.1).

5.2.3 Simulation Experiments

Simulations were conducted to compare performance of genotypes grown under well-watered and two drought conditions (Fig. 5.1), based on meteorological data registered in the greenhouse (Rebolledo et al. 2013). The well-watered condition (WW) was simulated by forcing soil moisture constantly at field capacity. The two drought simulations were differentiated as follows:

– 12d short, rapid dry-down (RDD) driven by the observed potential evapotranspiration (PET), resulting in a severe drought situation (32d simulations).
– 20d slow dry-down (SDD) driven by 60 % of observed PET, resulting in a moderate drought situation (40d simulations).

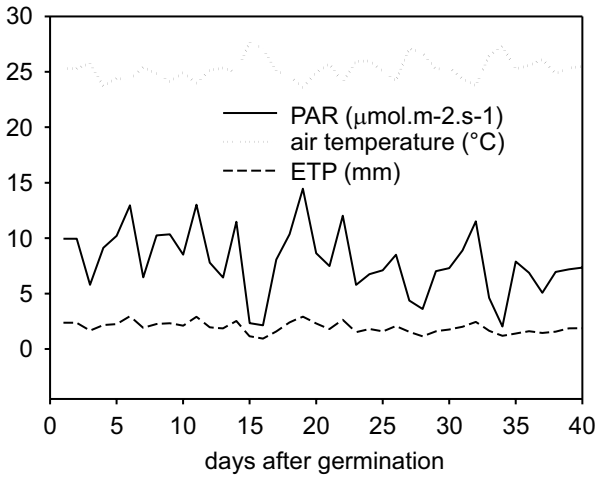


Fig. 5.1 Environmental conditions used for plant green shoot dry weight (GSDW) simulations with Ecomeristem presented in next figures (Data from greenhouse experiment presented by Rebolledo et al. (2013)). *PAR* photosynthetically active radiation, *ETP* evaporative demand. Simulations were performed from germination till 32 or 40 d after germination. For the moderate stress (40d simulation) the slow dry-down was simulated taking 40 % of actual ETP value presented in this Figure. For short severe dry-down, the real ETP value was used. For well-watered conditions, the soil is considered permanently at field capacity

For each simulation scenario, the behavior of each accession of the diversity panel was predicted. An additional set of simulations was conducted for the same scenarios using a panel of 9000 virtual genotypes (sensitivity analysis). The virtual genotypes represented combinations of parameter values within the ranges of observed on the real panel (Table 5.1). The parameter combinations for the virtual panel were generated with the FAST method (Saltelli et al. 1999).

5.2.4 Statistical Analyses

XLSTAT software was used for statistical analyses of simulated output variables and model parameters. Cluster analysis was performed with XLSTAT on phenotypes simulated for the genotypes on the diversity panel and for the phenotypes of the virtual panel generated by FAST. For this purpose, 32d and 40d simulations were considered separately. Cluster analysis was performed on the following simulation results: (i) final GSDW simulated under well watered (representing constitutive early vigor), and (ii) the reduction rate of final GSDW under each drought condition relative to WW (computed as [(100*GSDWww-GSDWd)/GSDWww], hereafter called GSDWred% or drought sensitivity). For the cluster analyses, simulation results obtained for the real panel and the virtual panel were combineds in a single data set (9136 individuals). A k-mean approach was first used to define 900 centroid genotypes (i.e., average of all genotypes in each of the 900 clusters generated by the k-mean approach) among the 9136 genotypes studied. Hierarchical ascendant classification (HAC) was then performed to cluster the 900 centroids in 9 classes, each class including potentially virtual or real accessions. Accordingly, nine classes each were created for early vigor (GSDW under WW), drought sensitivity (GSDWred%) under moderate drought (SDD, 40d simulations), and drought sensitivity under severe drought (RDD, 32d simulations). In each case, Principal Component Analysis (PCA) was performed to position clusters according to the parameter values of their respective genotypes.

5.3 Results

5.3.1 Predicted Shoot Dry Matter under Different Water Regimes

Two rice populations were studied, one real (diversity panel, 136 accessions) and one virtual (9000 “recombinants” of crop parameter values within the observed ranges, Table 5.1). For each real or virtual genotype, three phenotypes were generated corresponding to the three treatments. In the following we compare the six resulting populations of phenotypes, in terms of simulated GSDW at the end of the differential hydrological treatments.

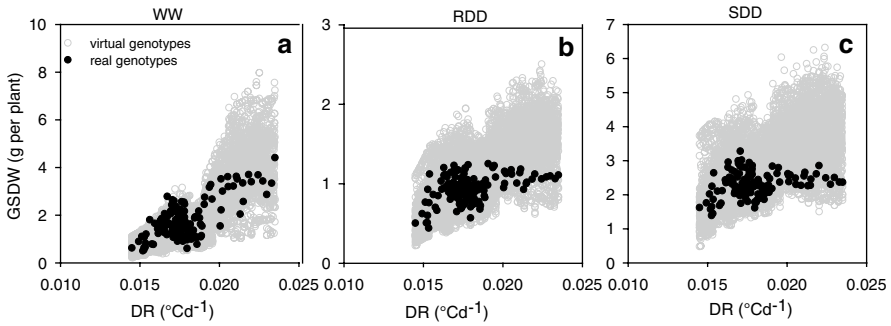


Fig. 5.2 Relationship between development rate (DR, °Cd^{-1}) and green-shoot dry weight (GDSW) accumulation simulated under three water conditions: (**a**) well-watered (WW), 32d simulation; (**b**) rapid severe 12d dry-down (RDD) ending at 32 days after germination; (**c**) slow moderate 20d dry-down (SDD) ending at 40 days after germination. *Black points*: 136 real genotypes constituting the diversity panel described in Table 5.1. *Grey points*: 9000 virtual genotypes generated within ranges of explored parameter values as described in Table 5.1

A previous study (Luquet et al. 2012b; Rebolledo et al. 2012a) described a positive correlation between GSDW and organogenetic development rate (DR, or phyllochron^{-1}; [°Cd^{-1}]) for real and virtual populations under WW treatment. This relationship was explored here for WW, RDD and SDD environments (Fig. 5.2). In all environments, simulated phenotypes of the real population represented a subset of the larger virtual population. The positive correlation between GSDW and DR was strong under WW for both real and virtual populations ($r = 0.72$ and 0.83, respectively; $p < 0.01$). Under drought, this correlation largely disappeared and was only visible as a trend among the genotypes having low DR (low-vigor types). For the real population, GSDW completely leveled off at DR > 0.017 °Cd^{-1}, corresponding to a phyllochron smaller than 59 °Cd, or 4.5d under the experimental conditions. These results indicate that differences in vigor (final GSDW under WW) were mostly not conserved under drought (RDD and SDD), but there was on average no penalty for high-vigor types in terms of absolute GSDW under drought.

Irrespective of the environments, GSDW predicted for a large part of the virtual population was greater than that for the real genotypes. This suggests that theoretically, within the model assumptions, a margin for better GSDW conservation under drought may exist.

The correlation between GSDW of drought treated vs. WW plants was positive. For both real and virtual populations, predicted GSDW of drought treated plants was either similar or inferior to that of WW (Fig. 5.3). For low-vigor types (low GSDW under WW), many genotypes showed no GSDW reduction under drought, whereas for high-vigor types there was generally a strong reduction of GSDW under drought, for both real and virtual genotypes. The virtual population, however, included a large proportion of cases that produced much greater GSDW than any of the real genotypes, regardless of treatment. Consequently, the virtual and real populations showed overall the same pattern of trade-offs between vigor and

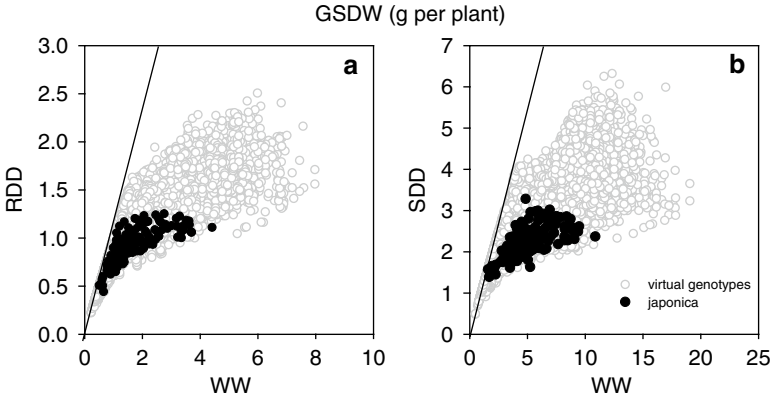


Fig. 5.3 Relationship between green-shoot dry weight (GSDW) simulated under well-watered WW and two drought treatments: (**a**) rapid severe 12d dry-down (RDD) ending at 32 days after germination; (**b**) slow moderate 20d dry-down (SDD) ending at 40 days after germination for the real and virtual panels. The 1:1 line represents genotypes performing equally well under both conditions

drought response, but the virtual population covered a much greater range of vigor (GSDW under WW).

5.3.2 Evaluation of Vigor and Drought Response by Cluster of Response Type

In order to facilitate the comparison among vigor and drought response types within the populations, the latter were divided into clusters having similar behavior. Clustering was based on two variables: (i) GSDW simulated under well-watered conditions that we will consider as the expression of early vigor, and (ii) the relative reduction of GSDW under drought as compared with WW, considered as being a measure of drought sensitivity.

For the RDD treatment, the nine clusters and their centroids were linearly lined up along a function resembling an inverse hyperbola or a logarithmic function (Fig. 5.4a). The two most vigorous clusters C8 and C9 did not contain real existing genotypes, whereas all other clusters did. The centroids of the three clusters having the poorest vigor and smallest drought sensitivity (C2, C4 and C6) were located within the range of values observed for real genotypes. The centroids for the other clusters were outside the domain observed for real genotypes and represented virtual genotypes having smaller drought sensitivity than similarly vigorous real genotypes. The nonlinear regression across the population of genotypes between GSDW and drought sensitivity can be expressed as (Eq. 5.1), showing a clear positive correlation between early vigor and drought sensitivity under such a drought type:

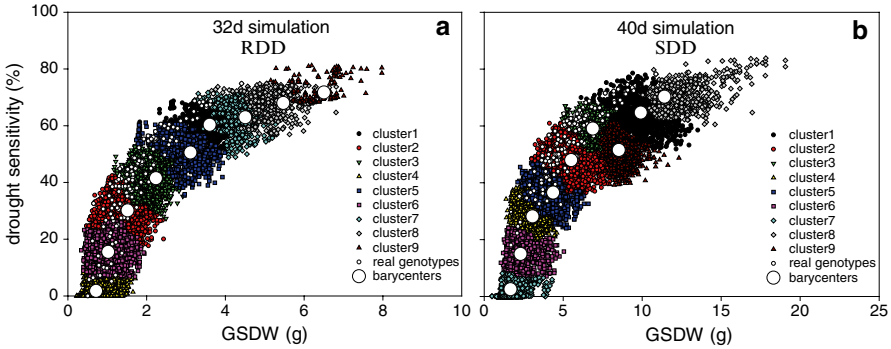


Fig. 5.4 Relationship between GSDW under well-watered conditions (WW) and its reduction rate by (**a**) a rapid dry-down event (RDD, 32d simulation ending by a 12d dry-down) and (**b**) by a longer and slow dry-down event (SDD, 40d simulation ending by a 20d dry-down). Clusters are distinguished by symbol color. *Large white circles* are the centroids for each cluster. *Small white symbols* represent the observed population

$$y = 17.8 + 29\ln(x) \left(r^2 = 0.82\right) \tag{5.1}$$

It is important to note that both virtual and real populations, and in particular the real population, occupied a remarkably narrow band on the drought sensitivity vs. vigor scatter plot (Fig. 5.4a). The trade-off between vigor and drought sensitivity (in terms of relative reduction in GSDW) was thus an inherent system property, or emergent feature from the simulations. The centroids of the clusters of virtual genotypes thereby showed a slightly smaller trade-off than did the real population.

For the SDD treatment, observations were similar as for RDD, but there was greater variation in drought sensitivity at given levels of vigor (Fig. 5.4b). This was true for both real and virtual populations, possibly indicating that there is inherently a greater margin for drought-vigor trade-offs when the stress progresses slowly (SDD) than when it is rapid and intense (RDD). However, the logarithmic correlations for drought sensitivity vs. vigor were similar (Eq. 5.2):

$$y = -5.7 + 29.7\ln(x) \left(r^2 = 0.8\right) \tag{5.2}$$

For SDD, six out of nine cluster centroids fell into regions occupied by real genotypes, and only centroids of high-vigor clusters C1, C8 and C9 fell into regions where only virtual genotypes were present.

5.3.3 *Principal Component Analysis for Model Parameter Values, Vigor and Drought Sensitivity*

Analysis of principal components explaining variation of genotypic model parameters, GSDW under WW (syn. vigor) and relative reduction of GSDW under drought (syn. drought sensitivity), conducted for the combined real and virtual populations,

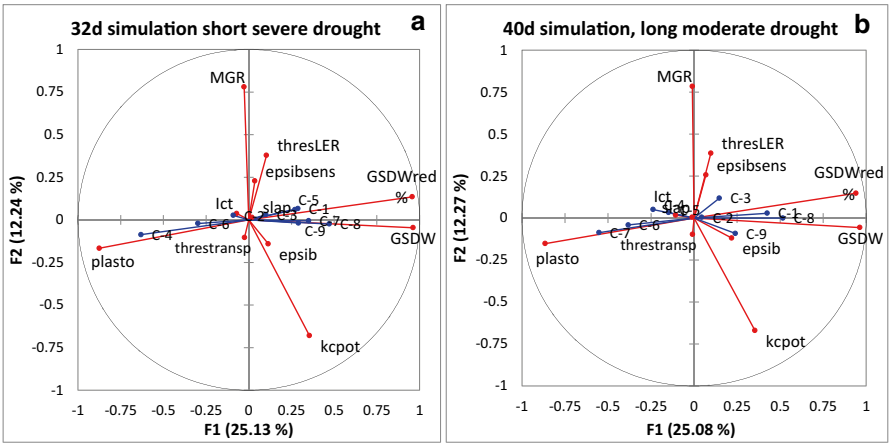


Fig. 5.5 Principal component analysis based on Ecomeristem model genotypic parameters (see Table 5.1 for their definition) and simulated green-shoot dry weight (GSDW g per plant) and its reduction rate under drought (GSDWred%); (**a**) for a rapid dry-down (RDD, 32d simulation ending by a 12d dry-down) and (**b**) for a longer and slower dry-down (SDD, 40d simulation ending by a 20d dry-down). Clusters (C-#) are represented by blue lines and markers

gave very similar patterns for RDD and SDD treatments (Fig. 5.5). On the PCA diagrams, model parameters (red) and clusters (blue) were depicted with different colors for distinction.

Vigor and drought sensitivity were closely associated with each other at the positive extreme of axis 1, confirming the strong trade-offs between vigor and drought effects. The model parameter *plasto*, which sets the potential duration of the phyllochron, was situated at the opposite, negative extreme of axis 1. Phyllochron is the reciprocal of organogenetic vigor, or development rate DR. Consequently, high DR (or low phyllochron) was associated with both vigor and drought sensitivity.

Strongly impacting parameters were also *MGR* (positively affecting organ size) and *Kcpot* (setting potential canopy transpiration) but their effect was opposite to each other and located mainly on axis 2, with little bearing on vigor and drought sensitivity. Among the lesser impacting parameters, *Ict* (negatively related to tillering ability) had an effect opposed to vigor. Thus, tillering ability contributed positively to vigor and drought sensitivity. *SLAp*, a parameter controlling the reduction rate of Specific Leaf Area (SLA) with increasing leaf position on the plant (the higher *SLAp*, the stronger SLA decrease from one position to the next), also impacted little on vigor and drought sensitivity. Its effect on both variables was null under RDD (Fig. 5.5a) and slightly positive for the milder SDD treatment (Fig. 5.5b). In general the impact of drought response parameters (*epsibsens*, *threstransp*, *thresLER*) was smaller compared with that of constitutive morphogenetic and physiological parameters. These three parameters participated in axis 2, meaning that they had little effect on vigor and drought sensitivity. Parameters *thresLER* and *epsibsens* were positively associated with *MGR* (conferring large leaves). Parameter *thresTransp* was negatively correlated with them.

Tables 5.2 and 5.3 exemplify changes in genotypic model parameter values (considered as traits) that would improve (as compared with real genotypes, and within the range of observed parameter values), (i) vigor at a given level of drought sensitivity, or (ii) drought sensitivity under a given level of constitutive vigor. The virtual population included phenotypes having much greater biomass under WW (constitutive vigor) than observed genotypes (Fig. 5.2a). Tables 5.2 and 5.3 suggest that for improved vigor (i.e., a shift to right from best existing genotypes in C1 to the centroids of C7 in Table 5.2 and C8 in Table 5.3) would involve traits such as greater leaf size (*MGR* increase) and smaller SLA, which translates into thicker leaves (*SLAp* increase). A slight increase of development rate (or shorter phyllochron) would also increase vigor, particularly for the longer simulation experiment (Table 5.3: −10 % for *plasto*). Surprisingly, no increase in tillering ability (decrease in parameter *Ict*) or light conversion efficiency (increase in parameter *epsib*) was suggested by the model to increase vigor.

Under drought conditions, the virtual population also included many cases of greater biomass than the best observed genotypes (Fig. 5.2b, c). According to Table 5.2, (RDD: rapid dry-down), a reduction of drought sensitivity by 23 % at a given level of vigor resulted from a lower drought sensitivity of both C assimilation (*epsibsens*, −11 %) and leaf expansion rate (*thresLER*, −31 %). This was associated with a strong reduction of tillering ability (*Ict*, +77 %), a 26 % increase in light conversion efficiency (*epsib*) and a reduction of potential canopy transpiration rate (*Kcpot*, −19 %). The more drought resistant plants would also have thicker leaves (*SLAp*, +16 %) but no change in potential leaf size (*MGR*) and number per culm (*plasto*).

Under a slower and milder dry-down (SDD, Table 5.3), a reduction of 26 % of drought sensitivity (downward-shift from existing genotypes in C1 toward the centroid of C9) involved a reduction in drought sensitivity of leaf extension rate (*thresLER* −49 %) and of light conversion efficiency (*epsibsens* −35 %), and a reduction in drought sensitivity of transpiration rate (*thresTransp*, −28 %). The more drought resistant plants would have slightly smaller tillering ability (*Ict*, +14 %) and lower potential leaf transpiration rate (*Kcpot*, −14 %). They would also have greater potential leaf size (*MGR* +35 %) and thicker leaves (*SLAp* +30 %).

5.4 Discussion

5.4.1 Critical Appreciation of the Conceptual Approach

The present study aimed at a plant model-based exploration of the trade-offs between drought sensitivity and vigor in terms of biomass production during vegetative growth. On this basis, an estimation of theoretical margins for increasing drought resistance or improving vigor, while reducing trade-offs, was attempted. Crop parameter values were thereby limited to vary within the ranges estimated for an observed diversity panel of rice, based on parameter optimization procedures.

Table 5.2 Model parameter values (see Table 5.1 for definitions), simulated green-shoot dry weight (GSDW) and its reduction rate (GSDWred) under RDD treatment (rapid, 12d dry-down) for: 12 real genotypes observed in Cluster 1 (C1), a virtual genotype being less drought sensitive (values for centroid of C5) and a virtual genotype being constitutively more vigorous (centroid of C7). For cluster behavior refer to Fig. 5.4a. The two bottom rows indicate changes in parameter values relative to the average of real genotypes in C1. Bold, underlined values indicate parameters for which this variation is superior to 10 %. Values in brackets are meaningless with respect to estimated change

		epsib	Ict	Kcpot	MGR	plasto	epsibsens	Slap	Thres LER	Thres Transp	GSDW (g)	GSDWred%
12 real genotypes in C1	Average	4.14	1.04	4.80	9.80	47.54	1.37	62.93	0.77	0.64	3.31	65.31
	SD	*0.17*	*0.23*	*0.24*	*1.79*	*3.12*	*0.13*	*7.91*	*0.15*	*0.14*	*0.23*	*2.62*
Centroid C5 (more tolerant)	–	5.21	1.85	3.87	10.24	50.96	1.22	73.12	0.53	0.63	3.11	50.54
Centroid C7 (greater vigor)	–	4.10	1.03	4.60	12.30	46.79	0.85	73.45	0.78	0.69	4.51	62.94
Change (%) to enhance drought tolerance (downward shift to C5)		**26**	**77**	**−19**	5	7	**−11**	**16**	**−31**	−2	−6	**−23**
Change (%) to enhance early vigor (right shift to C7)		−1	−2	−4	**25**	−2	(−38)	**17**	(1)	(8)	**36**	−4

Table 5.3 Model parameter values (see Table 5.1 for definitions), simulated green-shoot dry weight (GSDW) and its reduction rate (GSDWred) under SDD treatment (slow dry-down during 20d) for: 18 real genotypes observed in Cluster 1 (C1), a virtual genotype being less drought sensitive (values for centroid of C9) and a virtual genotype being constitutively more vigorous (centroid of C8). For cluster behavior refer to Fig. 5.4b. The two bottom rows indicate changes in parameter values relative to average of real genotypes in C1. Bold, underlined values indicate parameters for which this variation is superior to 10 %. Values in brackets are meaningless with respect to estimated change

		epsib	Ict	Kcpot	MGR	plasto	epsibsens	Slap	Thres LER	threstransp	GSDW (g)	GSDWred%
18 real genotypes in C1	Average	4.20	1.08	4.81	9.83	47.78	1.34	58.51	0.80	0.68	8.60	69.29
	SD	0.33	0.23	0.25	1.62	3.05	0.16	11.35	0.14	0.16	0.50	2.81
Centroid C9 (more tolerant)	–	3.98	1.23	4.15	13.22	50.70	0.68	76.29	0.52	0.49	8.52	51.51
Centroid C8 (greater vigor)	–	4.07	1.09	4.46	12.83	42.98	1.03	75.19	0.76	0.58	11.41	70.27
Change (%) to enhance drought tolerance (downward to C9)		−5	**14**	**−14**	**35**	6.10	**−49**	**30**	**−35**	**−28**	−0.91	**−26**
Change (%) to enhance early vigor (right shift to CL8)		−3	0.6	−7	**30**	**−10**	(−23)	**28**	(−4)	(−15)	**33**	2

By allowing parameter values (traits) to freely recombine within those limits, a larger virtual population was generated. The growth-environment conditions were characterized by a finite soil volume and water reserve to which the plants had full access and would deplete it as a function of simulated canopy transpiration and soil evaporation rate.

An integrative plant growth model, Ecomeristem, was used that recently demonstrated its suitability to simulate the diversity of rice seedling phenotypes under both well-watered (Luquet et al. 2012b) and drought conditions (Luquet et al. 2012a). Ecomeristem is a functional-structural plant model *(FSPM)* type (for more details of FSPM, see Chap. 2 by Xu and Buck-Sorlin in this book). The model simulates plant topology and morphogenetic processes and their phenotypic plasticity, in response to light and water resources. Resource availability results in plant internal nutritional status (supply/demand ratio of C or water) that regulates C source-sink regulation and morphogenesis depending on genotypic sensitivity. Genotypic model parameter values in Ecomeristem are generally process-based traits. They are therefore not static morphological traits but response traits that cannot be measured directly but are fitted to observed plant behavior by parameter optimization.

The challenge in this study was (i) to capture with model parameters existing genetic diversity; (ii) to explore potential genetic gain that can be possibly drawn from new combinations of traits, particularly with respect to the previously observed negative linkage between early vigor and drought tolerance within the existing genetic diversity (Rebolledo et al. 2012a, 2013); and (iii) to create ideotype concepts from phenomics resources as a potential strategic support for breeding (Bertin et al. 2010; Dingkuhn et al. 2007; Hammer et al. 2010).

The probable limitations to the validity of thus obtained information hinge on the accuracy of the model assumptions, the accuracy of the parameterization, and the hypothesis of free combinability of model parameter values within the observed ranges. These are all uncertain in our analysis. Although Ecomeristem has been validated and has proven its skill to capture phenotypic diversity in rice, particularly with respect to early vigor and drought response traits (Luquet et al. 2012a), none of the abovementioned uncertainties can be discounted in our context, because it carries model predictions into virtual genotypes. This study should therefore not be seen as predictive in an operational sense for breeding, but as a conceptual exploration of a largely new kind of crop model application.

Among the few precursor studies to our approach is the up-scaling of empirical, QTL-driven plant parameters for leaf extension rates under drought to the crop level using the crop model APSIM for maize (Chenu et al. 2009). That study, however, "implanted" the results of a reductionist, single-process study into the integrated crop-environment context (*APSIM*). However, the possible genetic linkages and pleiotropic effects were not taken into account, as the phenotypic and genetic studies were solely made on leaf elongation. In the present study a holistic approach was used, aiming at estimating in one procedure the values of genotypic parameters known to be most influential for water use, growth, architecture and development (Luquet et al. 2012a, b). Each parameter combination is thus supposed to represent and capture a genotype expressing its traits in interaction with other traits and with

the environment. The degree of uncertainty in such predictions may in fact be similar to that in the approach used by Chenu et al. (2009), but the present approach was able to integrate information on the genetic diversity observed for a much larger number of traits, including many traits constituting ideotypes.

5.4.2 *Experimental Design versus Type of Drought Resistance and Vigor*

The study focused on two drought treatments that were both progressive (dry-down using a finite water reserve) but differed in the rapidity of soil water depletion. In both cases, drought avoidance through deep root systems (Courtois et al. 2000) could not be expressed by the plants, and adaptations were thus related only to traits affecting plant water demand (in turn affected by leaf area growth), light interception and photosynthetic conversion, transpiration efficiency, and the control of various gas exchange and growth variables by soil water status (FTSW). The finite water resource in this scenario thereby generated an intrinsic trade-off between vigor and drought resistance. Since plant performance was evaluated based on biomass, the stress setup favored plants that are efficient water users and respond insensitively to water deficit within non-lethal ranges. The latter implies high levels of physiological tolerance. Under well-watered conditions, the biomass-based evaluation criterion favored radiation-use efficient (source) or rapidly developing and expanding (sink) plants, depending on whether growth was source or sink limited (Dingkuhn et al. 2007).

5.4.3 *Theoretical Scope for Improving Vigor*

Under WW conditions, the *in silico* exploration of phenotypes within the range of parameter values observed in the diversity panel indicated a theoretical margin for significant increases in biomass accumulation during vegetative growth (Fig. 5.2a; Tables 5.2 and 5.3). Traits that would contribute to this, according to the simulations, are largely on the assimilate demand side, such as larger leaves (*MGR* increase) and thicker leaves (*SLAp* increase) while maintaining at least the same developmental vigor (organogenetic vigor in terms of leaf number and tillering). In fact, Rebolledo et al. (2012a) already observed in the same diversity panel a negative linkage between leaf size and number, and growth appeared to be limited in many accessions by sink dynamics. Sink limitation was associated with accumulation of non-structural carbohydrates in the vegetative plant. This would suggest that there is still room to improve rice early vigor based on carbon sink rather than source related traits (Gibson et al. 2011; Ter Steege et al. 2005). Greater early vigor would thus not necessarily require greater photosynthetic potential. Although *a priori* surprising, sub-maximal biomass growth of a grain crop such as rice could be

a result of breeding and selection focused on grain yield and harvest index (Peng and Khush 2003). Early, excessive biomass accumulation might in fact be detrimental to harvest index in a high-yielding cereal (Dingkuhn et al. 1991).

5.4.4 *Theoretical Scope for Improving Drought Resistance*

Under drought conditions, the *in silico* exploration of phenotypes within the range of observed parameter values did indicate a theoretical scope for improvement of biomass growth (Fig. 5.2b, c). However, contrary to the case of vigor under non-limited water conditions, the adaptations for greater drought resistance were more physiological and less developmental (in the sense of organo- or morphogenetic processes). According to the results, a plant type being vigorous (in terms of biomass) under finite water resources would call for a combination of high radiation conversion efficiency and limited transpiration (resulting in high transpiration efficiency), and comparatively insensitive response of gas exchange rates to FTSW (necessarily associated with tissue level tolerance). Although model parameters were strictly used within the range of observed variation, there must be doubt regarding their independence from each other, and therefore their free combinability. Although evidence exists for genetic variation in transpiration efficiency (Rebolledo et al. 2013) and tissue-level drought tolerance in rice (Lilley et al. 1996; Robin et al. 2003), large gains in water economy can probably be expected only in partial or full implementation of C_4-type metabolic traits, as currently being engineered by the C_4-Rice consortium (von Caemmerer et al. 2012). A reliable exploration of the scope for improving such traits would require a model that is more mechanistic than Ecomeristem regarding the coupling of water- and carbon-use processes, and supported by phenotypic evidence for the diversity of the underlying physiological traits.

5.4.5 *Perspectives for Crop Modeling Support to Breeders and Geneticists*

Due to increasing knowledge on the genetic and environmental control of mechanisms underlying phenotype construction and behavior, theoretically, biological mechanistic modeling can in the future provide predictions of the phenotype to assist in the design and realization of crop genotypes. Limiting factors are not only lack of specific data and understanding of some key processes, but also appropriate tools that capture the complexity of the plant-environment system, namely the web of component interactions that give rise to emergent, new system properties at higher order (Baldazzi et al. 2013; Keurentjes et al. 2011). Ecomeristem is a model that attempts this with respect to phenotypic plasticity while being sufficiently sparsely parameterized to enable high-throughput fitting to phenotypic data on large

panels (Luquet et al. 2012a, b). The globally increasing efforts in phenotyping have become a resource not only for molecular genetic studies, which is their driving purpose, but also for predictive modeling of phenotypes and the characterization of their diversity. In the present study we benefitted from rice phenomics data that was generated for both gene discovery and modeling purposes. The GRiSP Global Rice Phenotyping Network initiated in 2011 (www.grisp.irri.org) will further increase the availability of data resources for such purposes.

The present study was conducted in conjunction with other researches not reported here, including the genotyping of the rice panel (genotyping by sequencing, GBS; McCouch et al. 2012) and genome-wide association studies (Courtois et al. 2013) for both directly observed traits and traits (model parameters) heuristically extracted from the data using Ecomeristem. It will thus be possible to implement known gene/allele effects in the plant model. At the same time, discovery of major genomic associations (QTLs) with model parameters will be a means to validate the biological relevance of the parameters, or conversely, identify those model parameters that have no significant biological basis in terms of genetic control (also see discussions in Chap. 9 by Yin et al.).

5.5 Conclusions

We attempted an *in silico* prediction of margins for genetic improvement of rice using a plant model specialized on morpho-physiological trait-trait interactions, Ecomeristem. The target was combined early vigor and drought resistance, based on virtual recombination of several traits (here syn. model parameters) within ranges of trait variation observed on a real panel of diverse rice genotypes.

Simulations of phenotypes under three water treatments indicated strong and similar trade-offs between constitutive vigor and drought resistance in both real and virtual populations. A substantial margin for potential genetic improvement of vigor with unchanged drought resistance was predicted, drawing chiefly from structural growth and development traits that would increase internal demand for assimilates (larger and thicker leaves, increased leaf appearance rates). Increased vigor would thereby not necessarily require greater photosynthetic potential per se. Conversely, improved drought resistance with unchanged constitutive vigor would require greater water economy (increased photosynthetic potential and limited water use, therefore greater transpiration efficiency) and greater tolerance of leaf extension and gas exchange processes to drought, while tillering ability should be limited in favor of larger and thicker leaves.

Although the results are physiologically plausible, the extrapolation involved in the study carries major uncertainties, namely with respect to (1) simplifications in the model, e.g. gas exchange and carbon assimilation processes; (2) the accuracy of model assumptions such as feedbacks of C source/sink ratios on growth and development; (3) errors in parameterization through error absorption among parameters during multi-fitting; and (4) the assumption of free combinability and additive expression

of traits. This study should therefore be seen as exploratory and in part conceptual. Upon further improvement, the methodologies may give rise to powerful tools for breeding support, particularly if genome-wide association studies can provide QTL effects that can drive phenotype predictions. Research on this is ongoing.

References

Araus JL, Slafer JA, Reynolds MP, Roy C (2002) Plant breeding and drought in C_3 cereals: what should we breed for? Ann Bot 89:925–940

Babu C, Shashidhar HE, Lilley JM, Thanh ND, Ray JD, Sadasivam S, Sarkarung S, O'Toole JC, Nguyen H-T (2001) Variation in root penetration ability, osmotic adjustment and dehydration tolerance among accessions of rice adapted to rainfed lowland and upland ecosystems. Plant Breed 120:233–238

Baldazzi V, Bertin N, de Jong H, Genard M (2013) Towards multiscale plant models: integrating cellular networks. Trends Plant Sci 17:728–736

Bertin N, Martre P, Génard M, Quilot B, Salon C (2010) Under what circumstances can process-based simulation models link genotype to phenotype for complex traits? Case study of fruit and grain quality traits. J Exp Bot 61:955–967

Brisson N, Gary C, Justes E, Roche R, Mary B, Ripoche D (2003) An overview of the crop model STICS. Eur J Agron 18:309–332

Chapman S, Cooper M, Podlich D, Hammer G (2003) Evaluating plant breeding strategies by simulating gene action and dryland environment effects. Agron J 95:99–113

Chenu K, Chapman SC, Tardieu F, McLean G, Welcker C, Hammer GL (2009) Simulating the yield impacts of organ-level quantitative trait loci associated with drought response in maize: a "gene-to-phenotype" modeling approach. Genetics 183:1507–1523

Cooper DW, Poldich DW, Micallef KP, Smith OS, Jensen NM, Chapman SC, Kruger NL (2002) Linking biophysical and genetics models to integrate physiology, molecular biology and plant breeding (Chapter 11). In: Kang MS (ed) Quantitative genetic, genomics and plant breeding, vol 1, CAB international. CAB Publishing, Wallingford, pp 143–166

Cooper M, Messina CD, Podlich D, Totir LR, Baumgarten A, Hausmann NJ, Wright D, Graham G (2014) Predicting the future of plant breeding: complementing empirical evaluation with genetic prediction. Crop Pasture Sci 65:311–336

Courtois B, McLaren G, Sinha PK, Prasad K, Yadav R, Shen L (2000) Mapping QTLs associated with drought avoidance in upland rice. Mol Breed 6:55–66

Courtois B, Audebert A, Dardou A, Roques S, Ghneim- Herrera T, Droc G, Frouin J, Rouan L, Gozé E, Kilian A, Ahmadi N, Dingkuhn M (2013) Genome-wide association mapping of root traits in a japonica rice panel. PLoS One 8, e78037

Dingkuhn M, de Vries FWT P, De Datta SK, van Laar HH (1991) Concepts for a new plant type for direct seeded flooded tropical rice. In: Direct seeded flooded rice in the tropics. International Rice Research Institute, Manila, pp 17–38

Dingkuhn M, Luquet D, Quilot B, Reffye PD (2005) Environmental and genetic control of morphogenesis in crops: towards models simulating phenotypic plasticity. Aust J Agr Res 56:1–14

Dingkuhn M, Luquet D, Clément-Vidal A, Tambour L, Kim HK, Song YH (2007) Is plant growth driven by sink regulation? Implications for crop models, phenotyping approaches and ideotypes. In: Struik PC, Spiertz HJ, van Laar HH (eds) Scale and complexity in plant systems research: gene-plant-crop relations, vol 21, Wageningen UR Frontis. Springer, Wageningen, pp 157–170

Gibson K, Park JS, Nagaia Y, Hwanga SK, Chod YC, Roh KH, Lee SM, Kim DH, Choie SB, Ito H, Edwards GE, Okita TW (2011) Exploiting leaf starch synthesis as a transient sink to elevate photosynthesis, plant productivity and yields. Plant Sci 181:275–281

Gu J, Yin X, Zhang C, Wang H, Struik PC (2014) Linking ecophysiological modelling with quantitative genetics to support marker-assisted crop design for improved yields of rice (*Oryza sativa*) under drought stress. Ann Bot 114:499–511

Hammer GL, Kropff MJ, Sinclair TR, Porter JR (2002) Future contributions of crop modelling – from heuristics and supporting decision making to understanding genetic regulation and aiding crop improvement. Eur J Agron 18:15–31

Hammer G, Sinclair TR, Chapman S, van Oosterom E (2004) On systems thinking, systems biology, and the in silico plant. Plant Physiol 134:909–911

Hammer GL, Chapman S, Van Oosterom E, Poldich DW (2005) Trait physiology and crop modelling as a framework to link phenotypic complexity to underlying genetic systems. Aust J Agr Res 56:947–960

Hammer G, van Oosterom E, McLean G, Chapman C, Broad I, Harland P, Muchow RC (2010) Adapting APSIM to model the physiology and genetics of complex adaptive traits in field crops. J Exp Bot 61:2185–2202

Jongdee B, Pantuwan G, Fukai S, Fischer K (2006) Improving drought tolerance in rainfed lowland rice: an example from Thailand. Agric Water Manage 80:225–240

Kamoshita A, Rodriguez R, Yamauchi A, Wade LJ (2004) Genotypic variation in response of rainfed lowland rice to prolonged drought and rewatering. Plant Prod Sci 7:406–420

Keurentjes JJB, Angenent GC, Dicke M, Martins Dos Santos VAP, Molenaar J, van der Putten WH, de Ruiter PC, Struik PC, Thomma B (2011) Redefining plant systems biology: from cell to ecosystem. Trends Plant Sci 16:183–190

Lilley JM, Ludlow MM, McCouch SR, O'Toole JC (1996) Locating QTL for osmotic adjustment and dehydration tolerance in rice. J Exp Bot 47:1427–1436

Luquet D, Dingkuhn M, Kim HK, Tambour L, Clément-Vidal A (2006) EcoMeristem, a model of morphogenesis and competition among sinks in rice. 1. Concept, validation and sensitivity analysis. Funct Plant Biol 33:309–323

Luquet D, Clément-Vidal A, This D, Fabre D, Sonderegger N, Dingkuhn M (2008) Orchestration of transpiration, growth and carbohydrate dynamics in rice during a dry-down cycle. Funct Plant Biol 35:689–704

Luquet D, Rebolledo MC and Soulié JC (2012a) Functional-structural plant modeling to support complex trait phenotyping: case of rice early vigour and drought tolerance using ecomeristem model. IEEE international symposium. 4 (PMA'12). In: Kang M, Dumont Y, Guo Y (eds) Plant growth modeling, simulation visualization and applications. IEEE, Shangai, pp. 270–277

Luquet D, Soulié JC, Rebolledo MC, Rouan L, Clément-Vidal A, Dingkuhn M (2012b) Developmental dynamics and early growth vigour in rice 2. Modelling genetic diversity using ecomeristem. J Agron Crop Sci 198:385–398

McCouch SR, McNally K, Wang W, Hamilton RS (2012) Genomics of gene banks: a case study in rice. Am J Bot 99:407–423

Namuco OS, Cairns JE, Johnson DE (2009) Investigating early vigour in upland rice (*Oryza sativa* L.): part I. Seedling growth and grain yield in competition with weeds. Field Crops Res 113:197–206

Nemoto K, Morita S, Baba T (1995) Shoot and root development in rice related to the phyllochron. Crop Sci 35:24–29

Pallas B, Clément-Vidal A, Rebolledo MC, Soulié JC, Luquet D (2013) Using plant growth modeling to analyze C source–sink relations under drought: inter- and intraspecific comparison. Front Plant Sci 4:437

Pantin F, Simonneau T, Rolland G, Dauzat M, Muller B (2011) Control of leaf expansion: a developmental switch from metabolics to hydraulics. Plant Physiol 156:803–815

Peng S, Khush GS (2003) Four decades of breeding for varietal improvement of irrigated lowland rice in the International Rice Research Institute. Plant Prod Sci 6:157–164

Quesnel G, Duboz R, Ramat E (2009) The virtual laboratory environment – an operational framework for multi-modelling, simulation and analysis of complex systems. Simul Model Pract Theory 17:641–653

Quilot B, Kervella J, Genard M, Lescourret F (2005) Analysing the genetic control of peach fruit quality through an ecophysiological model combined with a QTL approach. J Exp Bot 56:3083–3092

Quilot-Turion B, Ould-Sibi MM, Kadrani A, Hilgert N, Génard M, Lescourret F (2012) Optimization of parameters of the 'Virtual Fruit' model to design peach genotype for sustainable production systems. Eur J Agron 42:34–48

Rebolledo M-C, Dingkuhn M, Clement-Vidal A, Rouan L, Luquet D (2012a) Phenomics of rice early vigour and drought response: are sugar related and morphogenetic traits relevant? Rice 5:22

Rebolledo MC, Dingkuhn M, Péré P, McNally KL, Luquet D (2012b) Developmental dynamics and early growth vigour in rice. I. Relationship between development rate (1/phyllochron) and growth. J Agron Crop Sci 198:374–384

Rebolledo MC, Luquet D, Courtois B, Henry A, Soulié JC, Rouan L, Dingkuhn M (2013) Does early vigor occur in combination with drought tolerance and efficient water use in rice genotypes? Funct Plant Biol 40:582–594

Reymond M, Muller B, Tardieu F (2004) Dealing with the genotype x environment interaction via a modelling approach: a comparison of QTLs of maize leaf length or width with QTLs of model parameters. J Exp Bot 55:2461–2472

Richards R, Lukacs Z (2001) Seedling vigor in wheat–source of variation for genetic and agronomic improvement. Aust J Agr Res 53:41–50

Robin S, Pathan MS, Courtois B, Lafitte R, Carandang SLS, Amante M, Nguyen HT, Li Z (2003) Mapping osmotic adjustment in an advanced back-cross inbred population of rice. Theor Appl Genet 107:1288–1296

Saltelli A, Tarantola S, Chan K (1999) A quantitative, model independent method for global sensitivity analysis of model output. Technometrics 41:39–56

Sekhon S, Mebane WR Jr (2011) Genetic optimization using derivatives: the rgenoud package for R. J Stat Softw 42, 26 pp

Tardieu F, Granier C, Muller B (2011) Water deficit and growth. Co-ordinating processes without an orchestrator? Curr Opin Plant Biol 14:283–289

ter Steege MW, den Ouden FM, Lambers H, Stam P, Peeters AJM (2005) Genetic and physiological architecture of early vigor in *Aegilops tauschii*, the D-genome donor of hexaploid wheat. A quantitative trait loci analysis. Plant Physiol 139:1078–1094

Tisne S, Schmalenbach I, Reymond M, Dauzat M, Pervent M, Vile D, Granier C (2010) Keep on growing under drought: genetic and developmental bases of the response of rosette area using a recombinant inbred line population. Plant Cell Environ 33:1875–1887

von Caemmerer S, Quick WP, Furbank RT (2012) The development of C_4 rice: current progress and future challenges. Science 336:1671–1672

Xu L, Henke M, Zhu J, Kurth W, Buck-Sorlin G (2011) A functional structural model of rice linking quantitative genetic information with morphological development and physiological processes. Ann Bot 107:817–828

Yan HP, Kang MZ, Reffye PD, Dingkuhn M (2004) A dynamic, architectural plant model simulating resource-dependent growth. Ann Bot 93:591–602

Yin X, Struik PC (2010) Modelling the crop: from system dynamics to systems biology. J Exp Bot 61:2171–2183

Yin X, Kropff M, Stam P (1999) The role of ecophysiological models in QTL analysis: the example of specific leaf area in barley. Heredity 82:412–421

Yin X, Chasalow SD, Dourleijn CJ, Stam P, Kropff MJ (2000) Coupling estimated effects of QTLs for physiological traits to a crop growth model: predicting yield variation among recombinant inbred lines in barley. Heredity 85:539–549

Yin X, Struik PC, Kropff MJ (2004) Role of crop physiology in predicting gene-to-phenotype relationships. Trends Plant Sci 9:426–432

Chapter 6
Limited-Transpiration Trait for Increased Yield for Water-Limited Soybean: From Model to Phenotype to Cultivars

Thomas R. Sinclair, Jyostna M. Devi, and Thomas E. Carter Jr.

Abstract Soybean (*Glycine max* (L.) Merr.) is the most widely grown grain legume in the world due to its many uses in feed, food, and industrial products. However, soybean yield is particularly sensitive to soil water deficits, and seemingly, opportunities exist to increase yield by improving specific plant traits. One trait that has proven to be especially useful is the limited-transpiration trait in which water loss by the plants is constrained by the plant under high atmospheric vapor pressure deficit conditions. This chapter reviews the integrated studies at several levels and disciplines to identify the trait, develop some physiological and genetic understanding of the trait, apply classical breeding approaches to develop germplasm expressing the trait, and a simulation analysis across the USA to identify where, how often, and how much the trait in soybean will benefit farmers. The research on the limited-transpiration trait has now led to higher yielding commercial germplasm for water-deficit environments based on expression of the limited-transpiration trait. As often suggested but rarely put into practice, a multi-level, multi-faceted approach was applied in the study of the limited-transpiration trait to generate scientific understanding that was applied in crop breeding to generate higher yielding genotypes.

6.1 Introduction

Soybean (*Glycine max* (L.) Merr.) is grown globally to the greatest extent by far among grain legumes, with a total annual production of about 250 million tons (Sinclair and Vadez 2012). The high protein and oil contents of its seeds cause it to have many uses. Its protein is used to feed both livestock and humans, and its oil is

T.R. Sinclair (✉) • J.M. Devi
Crop Science Department, North Carolina State University, 27695-7620 Raleigh, NC, USA
e-mail: trsincla@ncsu.edu

T.E. Carter Jr.
Agricultural Research Service, US Department of Agriculture, 27695 Raleigh, NC, USA

X. Yin, P.C. Struik (eds.), *Crop Systems Biology*,
DOI 10.1007/978-3-319-20562-5_6

used in cooking. In addition, soybean oil has many industrial uses. Of course, a critical advantage of soybean is that it can express high rates of symbiotic nitrogen fixation alleviating the need for nitrogen fertilizer. Expanding consumption of meat and high prices of fossil fuel will result in continued increases in global production of soybean.

The increase in demand for soybean will likely require production in areas with uncertain rainfall patterns. Even now soybean is often subjected to sufficient soil drying that yields are constrained (Purcell and Specht 2004). One reason that soybean yield is especially sensitive to soil drying is the sensitivity of its symbiotic nitrogen fixation to even modest decreases in soil water content (Sinclair and Vadez 2012). Recently, we have given considerable attention to traits that result in early-season soil water conservation so that more water is available to complete seed fill under drought conditions, which occur more commonly late in the growing season. A specific trait that is especially promising in soybean is one in which transpiration rate is limited under high midday vapor pressure deficit. Partial restriction of transpiration rate under high vapor pressure deficit limits the rate of soil water use, allowing the crop to conserve water for sustaining physiological activity if late-season drought develops.

The objective of this chapter is to review the advances in developing the limited-transpiration trait in soybean. The approach to this trait generally followed the top-down approach leading to development of cultivars expressing the desired trait as originally proposed by Sinclair et al. (2004). In this approach, the sequence of steps undertaken include (1) initial exploration of the trait using a simulation model, (2) discovery of genotypic variation for the trait, (3) physiological studies on the nature of the trait, (4) genetic screen for the trait, and (5) development and assessment of cultivars with desired traits. Each of these steps is explored as essential components in progress in understanding the cropping system leading to yield increase.

6.2 Initial Model Exploration of Limited Transpiration Trait

The study of the limited-transpiration trait can be traced to a brainstorming session among Tom Sinclair, Graeme Hammer and Eric van Oosterom on one spring afternoon in 2004 under a tree at the University of Queensland in Brisbane, Australia. Dr. Sinclair was at the University of Queensland for two months with the rather ambiguous plan to "think" about plant traits that might improve crop yields under drought conditions. We ended up focusing on the idea that restricting transpiration under high atmospheric vapor pressure deficit (VPD) offered the possibility for conservation of soil water for use late in the growing season to support seed fill. We hypothesized that decreased transpiration rate under high VPD would have the double benefit of increasing transpiration water use efficiency, and conserving soil water for use later in the growing season.

We then undertook an initial assessment of the putative benefit of the limited-transpiration trait by simulating the development and growth of sorghum (*Sorghum*

bicolor L.) at four locations in Australia for which 115 years of weather had been assembled (Sinclair et al. 2005). The analysis was done using a comparatively simple, mechanistic sorghum/maize model (Sinclair et al. 1997; Sinclair and Muchow 2001). In this model, the increase in leaf area development is simulated daily as a function of temperature and constrained when soil water reaches low levels. The leaf area is used to calculate the daily growth of the crop by multiplying intercepted solar radiation by the radiation use efficiency. Radiation use efficiency was held constant except when it was decreased as soil water content reached low levels. Transpiration was calculated as a function of the crop growth, which was shown by Tanner and Sinclair (1983) to be an obligatory relationship with an essentially constant coefficient for each crop species. Seed growth was simulated as a linear increase in harvest index during seed fill.

The limited-transpiration trait was simulated by adapting the model from daily time step calculations to hourly time steps. Models to extrapolate daily minimum and maximum temperature, and solar radiation were used to obtain hourly estimates from daily weather input. Hourly values of vapor pressure deficit were calculated from the estimates of hourly temperature and the minimum daily temperature. The limited-transpiration trait was imposed by simply setting a maximum hourly transpiration rate. Whenever the initsial calculation of transpiration exceeded this limit, the transpiration rate was set equal to the limit. Also, in these cases hourly carbon accumulation was decreased to correspond to the decrease in transpiration rate. As a consequence, during the midday period the transpiration rate could be constant as illustrated in Fig. 6.1.

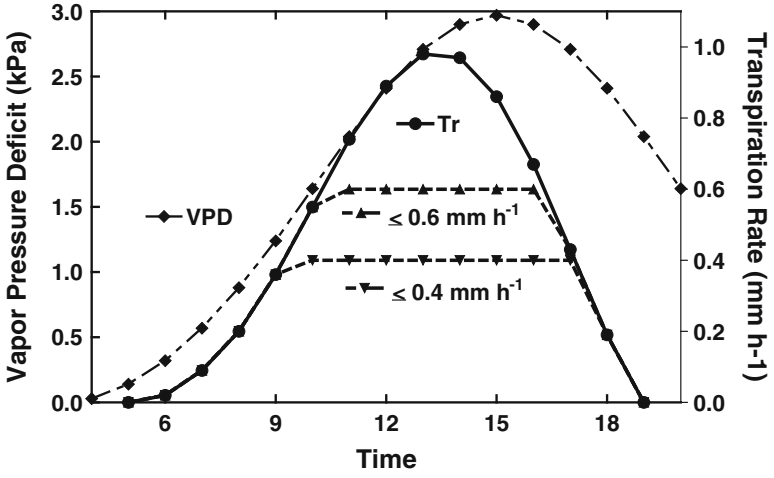


Fig. 6.1 Plot through the daily cycle of the transpiration rate of two limited-transpiration phenotypes (maximum transpiration rates of 0.4 or 0.6 mm h^{-1}) in which there is a constant transpiration rate once a maximum rate is reached (Sinclair et al. 2005). The limited-transpiration water loss pattern contrasts with the standard phenotype (*solid line*) in which there is no limitation on transpiration rate at high vapor pressure deficit. The vapor pressure deficit through the daily cycle is included as a reference

The results of the simulations for sorghum at the four locations in Australia were consistent. The mean yield increase over all seasons was in the range of 5 %. However more importantly, in growing seasons when yield was less than 450–500 g m^{-2} there was a yield increase in almost all growing seasons as a result of the transpiration-limited trait. The yield increase in these low-yielding seasons was generally around 10 %. In growing seasons above this threshold yield, yields were decreased due to the trait, but the decreases were generally quite small. Yield increases were simulated in about 75 % or more over all growing seasons. On balance, it was concluded that sorghum farmers in Australia would welcome the limited-transpiration trait since it resulted in yield increases in the economically threatening seasons of low yield, even if a small yield decrease was the price of this trait in the seasons of highest yield.

6.3 Studies of Limited-Transpiration Phenotype

Having shown in the sorghum model assessment that the limited-transpiration trait resulted in yield benefits in dry seasons, the next question was whether the idea also applies to soybean and if so, whether there is the possible expression of the trait in existing soybean germplasm. An experiment was done by one of the authors (TEC) that greatly narrowed the number of candidate genotypes that might express the limited-transpiration trait. A nursery of soybean genotypes was subjected to soil drying once full canopy had developed. During the drying cycle the onset of wilting for each genotype was observed. Two lines in particular were found to have delayed wilting with respect to all other lines: PI 416937 and PI 471938. Genotype PI 416937 was initially explored for several physiological traits for drought conditions but no specific trait was clearly identified to account for the delayed wilting (Sloane et al. 1990; Hudak and Patterson 1995).

Subsequently, the two ‘slow-wilting’ lines were investigated for the limited-transpiration trait (Fletcher et al. 2007). Genotype PI 416937 was found to express the desired limited-transpiration trait while PI 471938 did not (Fig. 6.2). In PI 416937, there was essentially a constant transpiration rate (Fig. 6.2c) at VPD greater than 2.1 kPa, which was an expression of the limited-transpiration trait explored in the sorghum model. Therefore, subsequent studies focused on PI 416937 in understanding the limited-transpiration trait. This genotype is a plant introduction from Japan with unknown parentage (Pantalone et al. 1999; Carter et al. 2003).

The hypothesis that was explored was that hydraulic conductance in the plants of PI 416937 was insufficient to allow water loss from leaves to be replenished under high VPD conditions. Three approaches were studied to determine if the hydraulic conductance in the leaves of PI 416937 was less than expressed in other genotypes (Sinclair et al. 2008). While all approaches indicated a low leaf hydraulic conductance in the leaves of PI 416937, the results for the temporal kinetics of rehydration of leaves were particularly interesting. The results showed that there were two distinct water compartments in soybean leaves, and both were much more

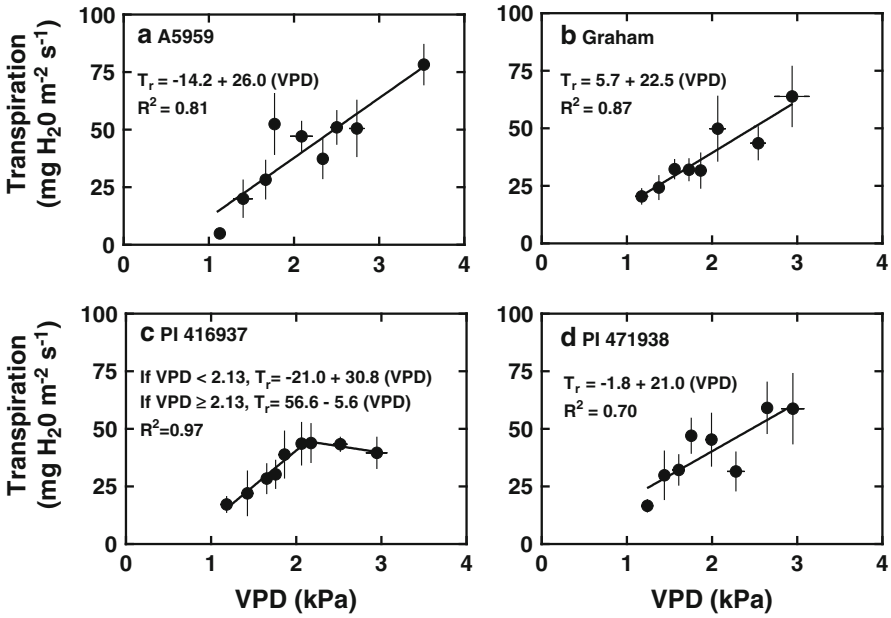


Fig. 6.2 Plot of transpiration rate for various vapor pressure deficits (VPD) of four soybean genotypes (Fletcher et al. 2007). Panel (**c**) presents the results for PI 416937 in which the transpiration rate was limited at a VPD threshold of 2.13 kPa. This result contrasts with other genotypes in which transpiration increased linearly over the whole range of tested VPD

slowly refilled in PI 416937 than in other genotypes. Of particular interest for the limited-transpiration trait, was the implication that hydraulic conductance between the xylem and the guard cells was low in PI 416937. A low hydraulic conductance in the leaves is likely to result in limited water flow to maintain turgor in guard cells for maximum pore opening.

Since no obvious unique features were observed in the leaf anatomy of PI 416937, the hypothesis was explored that the limited hydraulic conductance of PI 416937 may result from a unique population of aquaporins, i.e., water-transporting proteins in cell membranes. The population of aquaporins in PI 416937 was hypothesized to have lower capacity to transport water in the pathway from the xylem to guard cells. Sadok and Sinclair (2010a) subjected PI 416937 and a genotype without the limited-transpiration trait to two aquaporin inhibitors: mercury and silver. Transpiration of leaves of the two genotypes decreased equally when treated with mercury. However, when treated with silver ions the transpiration rate decreased substantially in the non-limited transpiration genotype while the transpiration rate of PI 416937 decreased only a small amount. Additionally, the comparative insensitivity of PI 416937 for decreasing transpiration rate to treatment with silver in comparison with many other soybean genotypes was documented by Sadok and Sinclair (2010b) (Fig. 6.3).

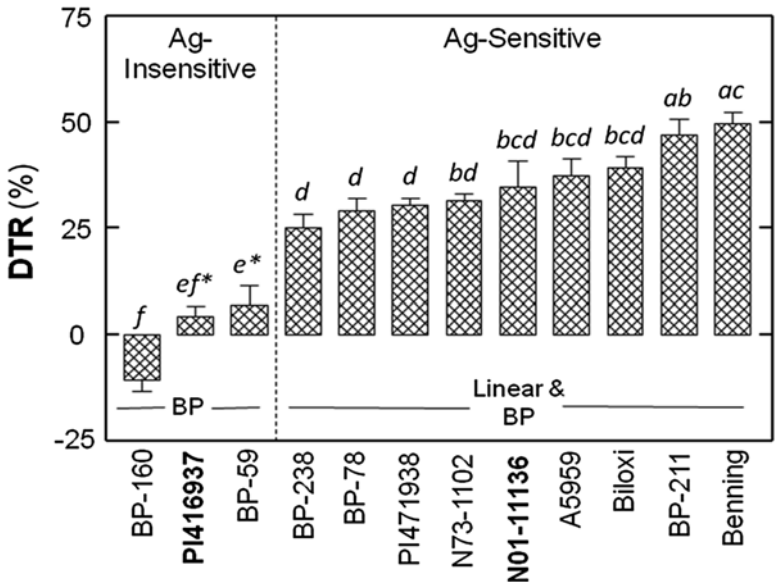


Fig. 6.3 Decrease in transpiration rate (DTR, expressed as percentage) of leaves of 12 soybean genotypes in response to the feeding with silver ion (Sadok and Sinclair 2010b). Line PI 416937 and two of its progeny lines (BP-160 and BP-59) were virtually insensitive to the silver treatment. These three lines were also found to express a breakpoint (BP) indicative of the limited-transpiration trait. The remaining lines had decreased transpiration rate with silver and these lines either expressed a BP or did not express a BP (i.e., a linear response to increasing vapor pressure deficit)

Sadok and Sinclair (2010b) hypothesized one explanation for the insensitivity of transpiration rate in PI 416937 to treatment with silver may be that it has fewer, or maybe none, of the aquaporins present in other genotypes that caused them to have decreased transpiration when treated with silver. The absence of the hypothesized silver-sensitive population of aquaporins in PI 416937 is consistent with the observation that the leaf hydraulic conductance of this genotype is less than that of other genotypes. That is, without the silver-sensitive aquaporins PI 416937 has restricted capacity for water movement to the guard cells resulting in limited transpiration rate under high VPD.

6.4 Studies of Limited-Transpiration Genotype

The possibility of sorting out the genetic expression of the limited-transpiration trait was explored in a recombinant inbred lines *(RILs)* population derived from the mating of PI 416937 and the cultivar Benning. While Benning also expressed a transpiration breakpoint in its response to increasing VPD (result not shown), the limitation on transpiration rate at high VPD was much less than in PI 416937. Sadok

and Sinclair (2009) found in a comparison of 22 RIL lines from the mating of PI 416937 × Benning that nine expressed the limited-transpiration trait at high VPD while thirteen did not. The breakpoint in the transpiration rate occurred at lower VPD (1.1–1.9 kPa) than for either parent.

A challenge in examining genotypic expression of the limited-transpiration trait is the limited capacity to measure directly the response of plants to a range of VPD. The possibility of using a screen based on transpiration response to silver treatment was explored since many genotypes (40+) could be tested in one day. Sadok and Sinclair (2010b) measured the response of five RILs to treatment with silver. The two RILs, which expressed the breakpoint in transpiration rate with increasing VPD, were also found to be insensitive to treatment with silver. The two RILs not expressing the breakpoint in transpiration with increasing VPD were quite sensitive to the silver treatment. However, one RIL that expressed a breakpoint was also sensitive to silver, indicating some ambiguity in interpreting the silver results.

Regardless of the specific mode of action of silver ions, the tests of Sadok and Sinclair (2010b) indicated that silver might be used as a rough screen of genotypes as a surrogate for direct measurements of the limited-transpiration trait. Therefore, a survey of the RIL population derived from PI 416937 × Benning was done for transpiration response to silver treatment of leaves (Carpentieri-Pipolo et al. 2012). A wide range in response was observed in the initial screen of 147 RILs. Forty-eight lines were selected from the extremes of the initial screen for retest for silver response and quantitative trait loci (QTL) analysis. Ultimately, four major QTLs were identified as being associated with the segregation of RILs for their silver response. The phenotypic variation explained by each of these QTLs was greater than 16 % and the total variation explained by the four QTLs was 87.5 %. Two of the loci appeared to be associated with PI 416937 and two with Benning.

A complementary approach to the identification of genes involved in the limited-transpiration trait is transcriptomics. Analysis of the expression patterns under high and low VPD reveals not just the genes induced by the high VPD but also the genes whose transcription is inhibited. Identification of several stress specific genes could help understand the physiological networks involved in stress responses.

Illumina Hiseq sequencing was performed on leaf tissues of three soybean genotypes: PI 416937, PI 471938, and Hutcheson, which is fast wilting with a linear increase in transpiration rate with increasing VPD. Differential expression was tested between exposure of plants to low VPD and high VPD under which the limited-transpiration trait would be expressed. Of the 49,408 annotated genes, only one gene differentially expressed in Hutcheson between exposure of low and high VPD and 22 in PI 471938. In contrast, PI 416937 differentially expressed 944 genes between exposure to low and high VPD conditions (Devi et al. 2015a).

With an objective to display differentially expressed genes onto pathways and to obtain an overview of genes affected in response to high VPD in PI 416937, the MapMan tool was used on the 944 genes in PI 416937 for which differential expression values were available. A total of 425 transcripts had up-regulated expression variation with log 2 fold changes from 2 to 9 and 519 with significant down-regulated expression with log 2 fold of −2 to −10. The overview map of Mapman

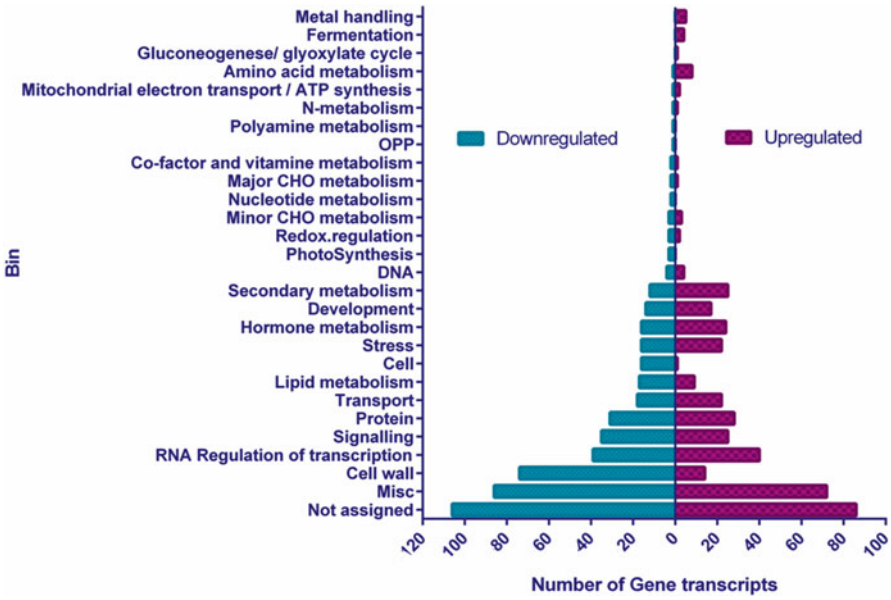


Fig. 6.4 Number of gene transcripts up or down regulated under high vapor pressure deficit conditions in soybean phenotype exhibiting limited transpiration trait. Annotated genes are categorized in to 28 functional classes (BINs) using MapMan

showed differentially expressed genes assigned into 28 classes of a total of 36 functional classes, referred as BINs (Thimm et al. 2004; Usadel et al. 2009). Of these classes, the majority of the genes were classified as unknown or not assigned, while the remaining were identified as belonging to known metabolic pathways (Fig. 6.4). This allowed exploration of gene categories that are activated during high VPD conditions and may have been involved in the processes associated with the limited-transpiration trait.

Based on the assigned genes to different BINs, an attempt was made to understand differentially expressed genes of key metabolic reactions that often modulate normal cellular functioning under the high VPD conditions. As a result, cell, cell wall and development, RNA, lipid metabolism, secondary metabolism, stress-related genes, protein, signaling and transport categories were analyzed in detail.

6.4.1 Cell, Cell Wall and Development

Changes in the magnitudes of cell and cell wall transcripts have been identified to play crucial roles in cellular metabolism. The genes related to cell metabolism (17) are mostly involved in the process of cell organization and are down regulated.

Genes coding cell wall (74) had decreased transcript abundance and the majority of those belong to cell wall degradation and modification (39) and some cell wall proteins are abundant in both up- and down-regulated categories. Some of the genes related to development mostly late embryogenesis abundant (LEA) proteins and storage proteins were found in both the up- and down-regulated categories.

6.4.2 Secondary, Hormone and Lipid Metabolism

Flavonoids and isoflavonoids are known to play a significant role in plant defense responses to pathogens (Dixon and Steele 1999; Uppalapati et al. 2009). Several genes related to secondary metabolism such as phenylpropanoids, flavanoids and simple phenols which are mostly over expressed were observed in response to high VPD. Several genes involved in phenylpropanoid metabolism, such as phenyl ammonia lyase, coumarate:CoA ligase, lignin biosynthesis, were observed in the study. All the aforementioned genes are in both over expressed and suppressed lists but commonly appeared in induction. Hormonal genes which are differentially over expressed are ethylene and jasmonate metabolism related genes. All abscisic acid, gibberellin, auxin and brassino-steroid related genes are repressed, that included some jasmonate metabolism related genes too. It is already well known that these hormones, especially ABA, are involved in stomatal regulation (Wang and Irving 2011). Lipid metabolism related genes include lipid degradation, sphingolipids, fatty acids (FA) synthesis and FA elongation and phospholipid synthesis which can be found both in expressed and inhibited category. FA may be an important determinant of responses of photosynthesis and stomatal conductance to environmental stresses such as high VPD (Poulson and Edwards 2002).

6.4.3 RNA Regulation of Transcription

Expression of limited transpiration responsive genes under high VPD environment was shown to be regulated by many transcription factors. Many genes (79) assigned to RNA regulation were identified. Genes coding for the zinc-finger family protein, MYB domain containing family, WRKY, ethylene response factor, bZIP were identified (Tran et al. 2004; Mochida et al. 2010). A high number of genes belonging to WRKY transcription factor and MYB domain category were found and they up-regulated under high VPD. Transcripts related to the basic helix-loop family were mostly decreased and the remaining transcripts like bZIP, zinc finger, and auxin/IAA were up and down regulated.

6.4.4 Protein, Signaling and Transport

Protein and signaling related genes were both induced and suppressed in high VPD samples. Protein genes are mostly involved in either post translational changes or degradation. Signaling genes include receptor kinases, leucine rich repeats and those involved in nutrient physiology. There were especially large numbers of the receptor kinases which are involved in improving plant performance under drought and also defense mechanism (Marshall et al. 2012). Several transport related transcripts are regulated under high VPD conditions and it was found that in PI 416937 most of the transcripts are induced (22).

Upregulated transcripts include nitrate, amino acids, ABC transporters, anion, cation, oligopeptides and phosphates. However, the majority of the differentially down-regulated transporters are major intrinsic proteins i.e., water channel proteins which probably could be the reason for the limited transpiration in PI 416937 under high VPD environments. In an aquaporin study in PI 416937 by Devi et al. (2015b), plasma membrane intrinsic proteins (PIP), especially PIP 2 were down regulated. PIPs and TIPs (Tonoplast Intrinsic Proteins) are said to play major roles in water transport (Maurel et al. 2008) and are responsive to different environmental conditions including VPD.

6.4.5 Stress Genes

Molecular responses to stress factors such as heat shock, anaerobiosis, plant pathogens, oxygen free radicals, heavy metals, water stress and chilling in plants have been assessed in various plant species (Matters and Scandalios 1986). Sixteen down-regulated and 22 up-regulated genes with biotic and abiotic stress-related annotations were grouped in to stress genes. Most of the biotic stress genes that are pathogen resistance responsive proteins were found to be more induced than suppressed, while abiotic stress showed an inverse pattern. The induced genes in the abiotic stress group include heat, drought and salt majorly involving heat shock proteins and dehydration responsive elements. The genes involved in the stress group are conserved and evidenced in earlier stress response studies of soybean (Le et al. 2012; Cal et al. 2013).

A clear trend in expression of all transcription factors together was not observed. However, overall, the differential regulation of many transcription factors under high VPD is similar to that seen with dehydration in soybean (Le et al. 2012).

6.5 Application of Limited-Transpiration Trait

While the initial simulations of the limited-transpiration trait with sorghum provided encouragement to study the trait, the value of the trait in developing cultivars for drought-tolerant lines and the possible benefit in soybean production needed to

be assessed. Two lines of evidence were developed in evaluating the practical benefit of the limited-transpiration trait. The first evidence involved the development of breeding lines that have superior performance under water-limited conditions. The breeding effort based on PI 416937 was initiated even before the results of the physiological studies were available. The second evidence was obtained for a detailed modeling of soybean production across the various environments of the USA to determine the amount and probability of yield increase that might be expected from the limited-transpiration trait.

6.5.1 Breeding Progress

Deliberate efforts to mitigate the impact of drought on soybean via breeding are relatively recent even though the problem has long been recognized by farmers and scientists (Carter 1989; Carter et al. 1999; Orf et al. 2004; Chen 2013). Until recently, the prevailing view among breeders was that yield data collected from drought stricken environments had little or no value because genetic repeatability or heritability of seed yield in these environments was thought to be (and often was) lower than in high-yielding irrigated counterparts. This view is exemplified by the practice of discarding yield trials from USDA regional testing (starting in the 1940s) whenever the average yield of the experiment was 170 g m^{-2} (25 bu/ac) or less. Minimum-stress environments were viewed as allowing expression of yield and greater separation of genotypic means. Sneller and Dombeck (1997) and Specht et al. (1986) offered arguments that generally supported this view. While they found some evidence for drought tolerance in the applied breeding pools in Arkansas and Nebraska, heritability and genetic variance for seed yield were generally greater in high-yielding irrigated environments.

A paradigm shift began with the discovery of the delayed wilting phenotype of PI 416937 (USDA 2012a) and PI 471938 (USDA 2012b) in the 1980s and 1990s (Sloane et al. 1990; Carter et al. 1999). Although genetic variation for seed yield was still regarded as greatest in high-yielding environments, the prospect of making agronomic yield improvements in drought-stressed environments gained substantial currency. Funding from the United Soybean Board (a non-profit farmer organization), starting in the mid 1990s, plus the identification of drought-prone field sites which had sufficient uniformity to support field breeding, set the stage for public breeders in Nebraska, Minnesota, North Carolina, Arkansas, and Georgia to begin the process of developing drought-tolerant cultivars.

One important field site identified for this work was North Carolina State University's Sandhills Research Station with deep uniform sandy soils and low water holding capacity. The station has moderate drought during August pod filling in 2 of 4 years, and extreme August drought in 1 of 4 years, on average. The USDA soybean breeding program in North Carolina began its drought breeding project in 1989 by hybridizing PI 416937 from Japan (identified in 1983 at a drought prone field site at Clinton, NC) with adapted USDA breeding line N77-144, known to be

an elite performer in high-yielding environments. F4-derived lines were evaluated under drought at the Sandhills station in replicated trials over 2 years, and eventually cultivar N7001 was released in 2003 from this effort (Carter et al. 2003). N7001 has good yielding ability and excellent overall yield stability and was the first USA cultivar that had in its pedigree as much as 50 % exotic pedigree since the 1950s. The USDA program has evaluated approximately 5000 yield plots annually at Sandhills research station since 1992. Several additional breeding lines have been developed, tested regionally, and shown to have high yield under drought (Devi et al. 2014). Some of these lines are being made available through Materials Transfer Agreements to major commercial seed companies and others for use as parental stock in their breeding programs. As such, this program is a major source of new genetic materials for the soybean industry and as genetic resources for physiological investigations into mechanisms of drought tolerance.

Using the cultivar N7001 as a parent, new cultivars N7002, N8001, and Woodruff were developed which are now among the top-yielding public cultivars in maturity groups VII and VIII (Carter et al. 2007, 2008). Both N7002 and N8001 have been high-yielding 'check' or control cultivars in USDA regional trials, in their respective maturity groups, for the past several years. A new breeding line, N05-7432, was developed more recently from the mating of N7002 (a derivative of PI 416937) with N98-7265 (a derivative of slow wilting PI 471938). This new breeding line is a top yielder in maturity group VIII (Gillen and Shelton 2012), surpassing check cultivar N8001 by 7 % ($p < 0.05$) averaged over more than 45 year-location combinations, which is quite large by breeding standards. The yield advantage of N05-7432 is quite stable, numerically out yielding N8001 in eight of the ten locations used in the multi-year testing trials. Further, expression of the limited-transpiration trait by N05-7432 has been documented (Devi et al. 2014). It also happens that N05-7432 is very tolerant to soil manganese deficiency (Masson 2014).

Line PI 416937 is perhaps the only exotic plant introduction being used as parental stock in USA at present which has the limited-transpiration trait. The full impact of this specific trait on improved agronomic drought response in its many progeny has not been ascertained as yet. Certainly, the ability of PI 416937 to conserve soil water to a greater extent than other soybean genotypes has been demonstrated in field experiments (King et al. 2009; Ries et al. 2012).

Parallel research in conjunction with field breeding has identified QTLs from PI 416937 for slow-wilting aquaporin response, prolific rooting, and aluminum tolerance in recent years, and all traits appear multi-genic in nature (Abdel-Haleem et al. 2011, 2012, 2013; Carpentieri-Pipolo et al. 2012). Current QTL research involves fine mapping of genetic markers for developing of factorial combinations of these QTL through breeding, so that their relation to the limited transpiration trait can be ascertained. In conjunction, the QTLs are being adapted to marker assisted selection in order to facilitate and enhance current field breeding programs aimed at improved drought tolerance.

6.5.2 *Model Assessment of Benefit in the USA*

Having investigated the physiology of the limited-transpiration trait and shown that the trait can be genetically transferred in a breeding program, key questions remain about where and by how much can soybean yield be expected to be impacted as a result of the trait. Considering the breadth of geographical area and environments in which soybean is grown in the USA, such an assessment can only be done by using simulations done over the range of locations and weather conditions likely to be experienced by soybean. Such an assessment using a mechanistic model has been done for soybean production in the USA (Sinclair et al. 2010).

The soybean model used in this study was originally presented by Sinclair (1986) and has been shown to be robust in several studies (Muchow and Sinclair 1986; Sinclair et al. 1987, 1992, 2007; Salado-Navarro and Sinclair 2009). The structure of the model is virtually the same as the sorghum model discussed earlier. The major modification was the simulation of nitrogen accumulation by the crop that reflected the high sensitivity of nitrogen fixation to soil drying. Loss of nitrogen fixation activity to soil drying caused an inhibition on leaf area development and ultimately less nitrogen in the plants to support seed growth. Again, the model was modified to run on an hourly basis to allow the limited-transpiration trait to be expressed in simulation during daytime hours under high VPD. In these simulations, the VPD breakpoint was assumed to occur at 2.0 kPa.

A key feature of the soybean simulations was the use of the GIS data base assembled by Pioneer DuPont (Loffler et al. 2005). In this data base, the areas in the US in which soybean is grown was segmented into 2655 blocks of 30 km × 30 km. In each block, 50 years of weather were developed to give daily minimum and maximum temperature, and precipitation. Solar radiation on each day was synthesized from temperature using the function developed by Bristow and Campbell (1984). In addition to weather information, the data base for each block included sowing date, maturity group, and available soil water storage.

Each set of simulations required the model to be run for each of the 50 years in each of the blocks for a total of more than 130,000 runs. The initial simulations were for a ‘standard’ soybean with a linearly increasing transpiration rate with increasing VPD. The limited-transpiration rate was simulated by imposing a VPD breakpoint at 2 kPa, above which there was no further increase in transpiration rate. For each year in each geographical block, the difference between the yield with the limited-transpiration trait and the standard soybean was calculated. A probability of yield increase was calculated for each block based on the fraction of years in which the yield increased as a result of the limited-transpiration trait. The results of the soybean simulations assessing the probability of yield increase as a result of the limited-transpiration trait are shown in Fig. 6.5a. The probability of yield increase was 80 % or greater in virtually all areas of the USA. The only places where there was not a high probability of yield increase were on the coasts in the southeast, the very

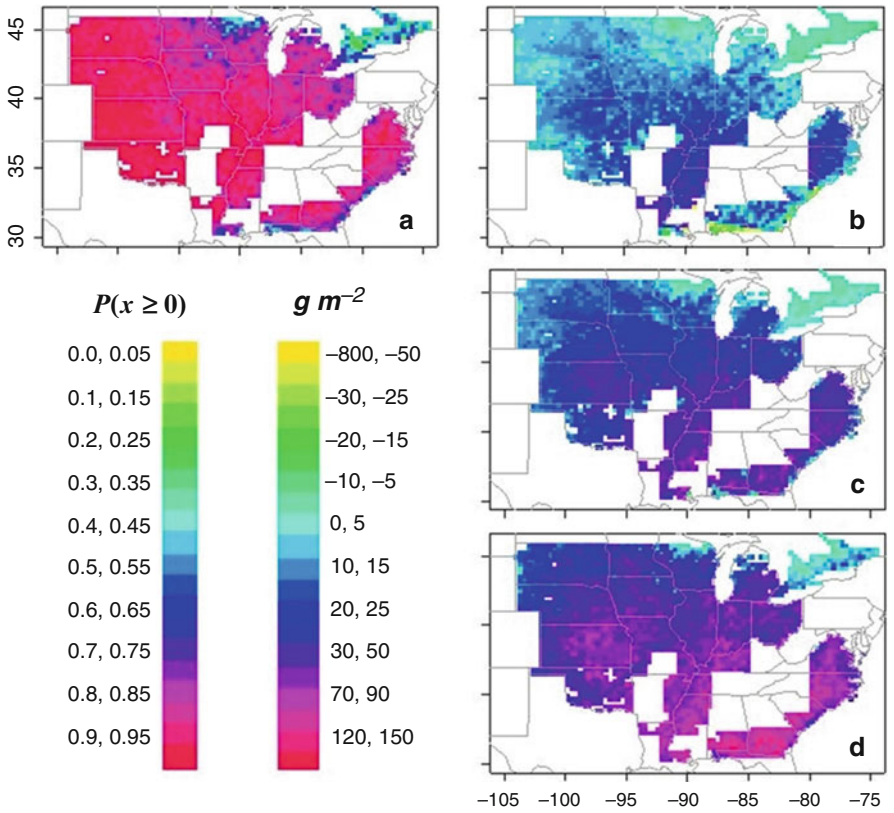


Fig. 6.5 Simulation results of soybean grain yield with limited-transpiration trait for 50 years in 30×30 km blocks across the USA (Sinclair et al. 2010). Panel (**a**) presents the probability of yield increase over the standard soybean in each block. Panels (**b**), (**c**), and (**d**) present the 75 % (wet), medium, and 25 % (dry) percentile yield increase (in g m^{-2}) over the standard soybean within each block

northern blocks, and a few scattered blocks in Iowa and Minnesota. Overall, the probability results show the limited-transpiration trait to be beneficial in nearly all the soybean production areas in the USA.

In addition to calculating the probability of yield increase, within each block the yield difference of each growing season was ranked from lowest to highest. This ranking gave the distribution yield from which specific percentile rankings could be compared. In Fig. 6.5b–d are shown the yield difference for each block for the 75 % (wet years), median, and 25 % (dry years) percentile, respectively. In the 25 % percentile, or drier years, the model predicted increases of about 90 g m^{-2} (dry grain weight) in many locations especially in the southern areas. The southeast, mid-south, and Kansas showed the greatest benefit of the limited-transpiration trait. The remainder of the country had yield increases in the 25 % percentile years in the range of 30–90 g m^{-2}.

The benefit of the limited-transpiration trait was also positive in most areas in the USA in the median year. In these years the yield benefit was simulated to be in the range of roughly 20–70 g m^{-2}. Even in the 75 % percentile year, the limited-transpiration trait was simulated to be beneficial although the yield increase tended to be in the range of 0–25 g m^{-2}. The largest benefit in the 75 % percentile year was in the North Carolina & Virginia, the Mississippi Delta region, and a belt from southern Indiana to Kansas. While not shown in these maps, the yield decrease simulated in the wettest years was small.

6.6 Conclusions

This study of the limited-transpiration trait in soybean illustrates the benefit of a systematic approach involving approaches and disciplines of study. Simulation studies were done initially to assess the potential value of the trait. Potential germplasm sources for the trait were evaluated in field screens and candidate lines were subjected to detailed studies of transpiration response across various levels of VPD. Physiological investigations of the trait led to hypotheses for investigation of the limited-transpiration trait and its potential physiological explanation. Finally, breeding has progressed to the release of soybean lines that express the limited-transpiration and have increased grain yields under dry conditions. Another series of simulations were done specifically for the production areas for soybean in the USA. These simulations indicated where, how often, and how much the trait in soybean will benefit farmers. The results of the studies on the limited-transpiration trait were advanced due to a comprehensive research program that involved investigations in several disciplines, under field and controlled conditions, and at several levels of sophistication. As often suggested but rarely put into practice, a multi-level approach in the study of the limited-transpiration trait involving a multi-faceted research effort resulted in progress in scientific understanding leading to benefit for farmers.

References

Abdel-Haleem H, Lee GJ, Boerma HR (2011) Identification of QTL for increased fibrous roots in soybean. Theor Appl Genet 122:935–946

Abdel-Haleem H, Carter TE Jr, Purcell LC, King CA, Ries LL, Chen P, Schapaugh W Jr, Sinclair TR, Boerma HR (2012) Mapping quantitative loci for canopy wilting trait in soybean (*Glycine max* L. Merr.). Theor Appl Genet 125:837–846

Abdel-Haleem H, Carter TE Jr, Rufty TE, Boerma HR, Li Z (2013) Quantitative trait loci controlling aluminum tolerance in soybean: candidate gene and SNP marker discovery. Mol Breed 33:851–862

Bristow KL, Campbell GS (1984) On the relationship between incoming solar radiation and daily maximum and minimum temperature. Agric For Meteorol 31:159–166

Cal AJ, Liu D, Mauleon R, Hsing Y-IC, Serraj R (2013) Transcriptome profiling of leaf elongation zone under drought in contrasting rice cultivars. PLoS One 8, e54537
Carpentieri-Pipolo V, Pipolo AE, Abdel-Haleem H, Boerma HR, Sinclair TR (2012) Identification of QTLs associated with limited leaf hydraulic conductance in soybean. Euphytica 186:679–686
Carter TE Jr (1989) Breeding for drought tolerance in soybean: where do we stand? In: Pascale AJ (ed) Proceedings of the world soybean conference IV, Asociacion Argentina de la Soja, Buenos Aires, 1856, pp 1001–1008
Carter TE Jr, De Souza PI, Purcell LC (1999) Recent advances in breeding for drought and aluminum resistance in soybean. pp 106–125. In: Kauffman H (ed) Proceedings of the world soybean conference VI Chicago, IL. 4–7 Aug. 1999. Superior Print, Champaign, pp 542
Carter TE Jr, Burton JW, Bowman DT, Cui Z, Zhou X, Villagarcia MR, Fountain MO, Niewoehner AS (2003) Registration of 'N7001' soybean. Crop Sci 43:1126–1127
Carter TE Jr, Burton JW, Fountain MO, Rzewnicki PE, Villagarcia MR, Bowman DT (2007) Registration of 'N7002' soybean. J Plant Reg 1(2):93–94
Carter TE Jr, Burton JW, Fountain MO, Rzewnicki PE, Villagarcia MR, Bowman DT (2008) Registration of N8001 soybean. J Plant Regist 1(2):23–23
Chen P (2013) Breeding for drought tolerance in soybean. In: World soynbean conference Durbin, South Africa, Jan 2013
Devi JM, Sinclair TR, Chen P, Carter TE (2014) Evaluation of elite southern maturity soybean breeding lines for drought-tolerant traits. Agron J 106:1947–1954
Devi JM, Sinclair TR, Taliercio E (2015a) Comparisons of the effects of vapor pressure deficit on gene expression among fast wilting and slow wilting soybean leaves. PloS One 10(10), e0139134
Devi JM, Sinclair TR, Taliercio E (2015b) Silver and zinc inhibitors influence transpiration rate and aquaporin transcript levels in intact soybean plants. Env Exp Bot (in press)
Dixon RA, Steele CL (1999) Flavonoids and isoflavonoids – a gold mine for metabolic engineering. Trends Plant Sci 4:394–400
Fletcher AL, Sinclair TR, Allen LH Jr (2007) Transpiration responses to vapor pressure deficit in well watered 'slow-wilting' and commercial soybean. Environ Exp Bot 61:145–151
Gillen AM, Shelton GW (2012) Uniform soybean tests – Southern States, 2011. USDA-ARS, Stoneville, MS. Available at http://www.ars.usda.gov/Business/Business.htm?modecode=64-02-15-00&docid=4357&page=3. Verified 29 Nov 2013
Hudak CM, Patterson RP (1995) Vegetative growth analysis of a drought-resistant soybean plant introduction. Crop Sci 35:464–471
King CA, Purcell LC, Brye KR (2009) Differential wilting among soybean genotypes in response to water deficit. Crop Sci 49:290–298
Le DT, Nishiyama R, Watanabe Y, Tanaka M, Seki M, Yamaguchi-Shinozaki K, Shonozaki K, Tran LS (2012) Differential gene expression in soybean leaf tissues at late developmental stages under drought stress revealed by genome-wide transcriptome analysis. PLOS One 7, e49522
Loffler CM, Wei J, Fast T, Gogerty J, Langton S, Bergman M, Merrill B, Cooper M (2005) Classification of maize environments using crop simulations and geographical information systems. Crop Sci 45:1708–1716
Masson R (2014) Heritability and selective genotyping QTL analysis of resistance to manganese deficiency derived from soybean breeding line SC97-1821. North Carolina State University, Raleigh
Marshall A, Aalen RB, Audenaert D, Beeckman T, Broadley MR, Butenko MA, Caño-Delgado AI, de Vries S, Dresselhaus T, Felix G (2012) Tackling drought stress: receptor-like kinases present new approaches. Plant Cell Online 24:2262–2278
Matters GL, Scandalios JG (1986) Changes in plant gene expression during stress. Develop Genetics 7:167–175

Maurel C, Verdoucq L, Luu DT, Santoni V (2008) Plant aquaporins: membrane channels with multiple integrated functions. Ann Rev Plant Biol 59:595–624

Mochida K, Yoshida T, Sakurai T, Yamaguchi-Shinozaki K, Shinozaki K, Tran LSP (2010) LegumeTFDB: an integrative database of Glycine max, Lotus japonicus and Medicago truncatula transcription factors. Bioinformatics 26:290–291

Muchow RC, Sinclair TR (1986) Water and nitrogen limitations in soybean grain production. II. Field and model analysis. Field Crops Res 15:143–156

Orf JH, Hutton SF, Carter TE Jr (2004) Breeding for soybean drought tolerance: update on U.S. experience. pp 260–266. In: Moscardi F, Hoffmann-Campo CB, Ferreira Saraiva O, Galerani PR, Krzyzanowski FC, Carrao-Panizzi MC (ed) Proceedings of the world soybean conference VII, Foz do Iguassu, PR, Brazil, February 29–March 5 2004. Embrapa Soybean, Londrina, pp 1344

Pantalone VR, Rebetzke GJ, Burton JW, Carter TE Jr, Israel DW (1999) Soybean PI 416937 root system contributes to biomass accumulation in reciprocal grafts. Agron J 91:840–844

Poulson ME, Edwards GE (2002) Photosynthesis is limited at high leaf to air vapor pressure deficit in a mutant of Arabidopsis thaliana that lacks trienoic fatty acids. Photosyn Res 72:55–63

Purcell L, Specht JE (2004) Physiological traits for ameliorating drought stress. In: Boerma HR, Specht JE (eds) Soybeans: improvement, production, and uses. American Society of Agronomy, Madison, pp 569–620

Ries LL, Purcell LC, Carter TE Jr, Edwards JT, King CA (2012) Physiological traits contributing to differential canopy wilting in soybean under drought. Crop Sci 52:272–281

Sadok V, Sinclair TR (2009) Genetic variability of transpiration response to vapor pressure deficit among soybean (Glycine mas [L.] Merr.) genotypes from a recombinant inbred line population. Field Crops Res 113:156–160

Sadok W, Sinclair TR (2010a) Transpiration response of 'slow-wilting' and commercial soybean (*Glycine max* (L.) Merr.) genotypes to three aquaporin inhibitors. J Exp Bot 61:821–829

Sadok W, Sinclair TR (2010b) Genetic variability of transpiration response of soybean [*Glycine max* (L.) Merr.] shoots to leaf hydraulic conductance inhibitor AgNO3. Crop Sci 50:1423–1430

Salado-Navarro LR, Sinclair TR (2009) Crop rotations in Argentina: analysis of water balance and yield using crop models. Agr Syst 102:11–16

Sinclair TR (1986) Water and nitrogen limitations in soybean grain production. I. Model development. Field Crops Res 15:125–141

Sinclair TR, Muchow RC (2001) System analysis of plant traits to increase grain yield on limited water supplies. Agron J 93:263–270

Sinclair TR, Vadez V (2012) The future of grain legumes in cropping systems. Crop Pasture Sci 63:501–512

Sinclair TR, Muchow RC, Ludlow MM, Leach GJ, Lawn RJ, Foale MA (1987) Field and model analysis of the effect of water deficits on carbon and nitrogen accumulation by soybean, cowpea, and black gram. Field Crops Res 17:121–140

Sinclair TR, Salado-Navarro LR, Morandi EN, Bodrero ML, Martignone RA (1992) Soybean yield in Argentina in response to weather variation among cropping seasons. Field Crops Res 30:1–11

Sinclair TR, Muchow RC, Monteith JL (1997) Model analysis of sorghum response to nitrogen in subtropical and tropical environments. Agron J 89:201–207

Sinclair TR, Purcell LC, Sneller CH (2004) Crop transformation and the challenge to increase yield potential. Trends Plant Sci 9:70–75

Sinclair TR, Hammer GL, van Oosterom EJ (2005) Potential yield and water-use efficiency benefits in sorghum from limited maximum-transpiration rate. Funct Plant Biol 32:945–952

Sinclair TR, Salado-Navarro LR, Salas G, Purcell LC (2007) Soybean yields and soil water status in Argentina: simulation analysis. Agr Syst 94:471–477

Sinclair TR, Zwieniecki MA, Holbrook NM (2008) Low leaf hydraulic conductance associated with drought tolerance in soybean. Physiol Plant 132:446–451

Sinclair TR, Messina CD, Beatty A, Samples M (2010) Assessment across the United States of the benefits of altered soybean drought traits. Agron J 102:475–482

Sloane RJ, Patterson RP, Carter TE Jr (1990) Field drought tolerance of a soybean plant introduction. Crop Sci 30:118–123

Sneller CH, Dombeck D (1997) Use of irrigation in selection for soybean yield potential under drought. Crop Sci 37:1141–1147

Specht JE, Williams JH, Weidenbenner CJ (1986) Differential responses of soybean genotypes subjected to a seasonal soil water gradient. Crop Sci 26:922–934

Tanner CB, Sinclair TR (1983) Efficient water use in crop production: research or re-search? In: Taylor HM, Jordan WR, Sinclair TR (eds) Limitations to efficient water use in crop production. American Society of Agronomy, Madison, pp 1–27

Thimm O, Blasing O, Gibon Y, Nagel A, Meyer S, Kruger P, Selbig J, Muller LA, Rhee SY, Stitt M (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J 37:914–939

Tran LS, Nakashima K, Sakuma Y, Simpson SD, Fujita Y, Maruyama K, Fujita M, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2004) Isolation and functional analysis of *Arabidopsis* stress-inducible NAC transcription factors that bind to a drought-responsive *cis-element* in the early responsive to dehydration stress promoter. Plant Cell 16:2481–2498

Uppalapati SR, Marek SM, Lee HK, Nakashima J, Tang YH, Sledge MK, Dixon RA, Mysore KS (2009) Global gene expression profiling during Medicago truncatula–Phymatotrichopsis omnivore interaction reveals a role for jasmonic acid, ethylene, and the flavonoid pathway in disease development. Molecular Plant Microbe Interact 22:7–17

USDA-ARS National Genetic Resources Program (USDA-ARS NGRP) (2012a) Glycine max (L.) Merr. PI 416937 Germplasm Resources Information Network (GRN). National Germplasm Resources Laboratory, Beltsville. Available at http://www.ars-grin.gov/cgi-bin/npgs/acc/display.pl?1314868. Verified 28 Sept 2013

USDA-ARS National Genetic Resources Program (USDA-ARS NGRP) (2012b) Glycine max (L.) Merr. PI 471938 Germplasm Resources Information Network (GRIN). National Germplasm Resources Laboratory, Beltsville. Available at http://www.ars-grin.gov/cgi-bin/npgs/acc/display.pl?1366874. Verified 28 Sept 2013

Usadel B, Poree F, Nagel A, Lohse M, Czedik-Eysenberg A, Stitt M (2009) A guide to using MapMan to visualize and compare omics data in plants: a case study in the crop species, maize. Plant Cell Environ 32:1211–1229

Wang YH, Irving HR (2011) Developing a model of plant hormone interactions. Plant Signal Behav 6:494–500

Chapter 7
Molecular Breeding for Complex Adaptive Traits: How Integrating Crop Ecophysiology and Modelling Can Enhance Efficiency

Graeme Hammer, Charlie Messina, Erik van Oosterom, Scott Chapman, Vijaya Singh, Andrew Borrell, David Jordan, and Mark Cooper

Abstract Progress in crop improvement is limited by the ability to identify favourable combinations of genotypes (G) and management practices (M) in relevant target environments (E) given the resources available to search among the myriad of possible combinations. To underpin yield advance we require prediction of phenotype based on genotype. In plant breeding, traditional phenotypic selection methods have involved measuring phenotypic performance of large segregating populations in multi-environment trials and applying rigorous statistical procedures based on quantitative genetic theory to identify superior individuals. Recent developments in the ability to inexpensively and densely map/sequence genomes have facilitated a shift from the level of the individual (genotype) to the level of the genomic region. Molecular breeding strategies using genome wide prediction and genomic selection approaches have developed rapidly. However, their applicability to complex traits remains constrained by gene-gene and gene-environment interactions, which restrict the predictive power of associations of genomic regions with phenotypic responses. Here it is argued that crop ecophysiology and functional whole plant modelling can provide an effective link between molecular and organism scales and enhance molecular breeding by adding value to genetic prediction approaches. A physiological

G. Hammer (✉) • E. van Oosterom • V. Singh
Centre for Plant Science, Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Brisbane, Qld 4072, Australia
e-mail: g.hammer@uq.edu.au

C. Messina • M. Cooper
DuPont-Pioneer, 7250 NW 62nd Avenue, PO Box 552, Johnston, IA 50131, USA

S. Chapman
CSIRO Plant Industry and Climate Adaptation Flagship, Queensland Bioscience Precinct, 306 Carmody Road, St Lucia QLD 4067, Australia

A. Borrell • D. Jordan
Centre for Plant Science, Queensland Alliance for Agriculture and Food Innovation, Hermitage Research Facility, The University of Queensland, 604 Yangan Road, Warwick, Qld 4370, Australia

X. Yin, P.C. Struik (eds.), *Crop Systems Biology*,
DOI 10.1007/978-3-319-20562-5_7

framework that facilitates dissection and modelling of complex traits can inform phenotyping methods for marker/gene detection and underpin prediction of likely phenotypic consequences of trait and genetic variation in target environments. This approach holds considerable promise for more effectively linking genotype to phenotype for complex adaptive traits. Specific examples focused on drought adaptation are presented to highlight the concepts.

Keywords Genotype-to-phenotype • Phenotypic prediction • Trait physiology • QTL • Functional genomics • Crop improvement

7.1 Introduction

Progress in crop improvement is limited by the ability to identify favourable combinations of genotypes (G) and management practices (M) in the relevant target environments (E) given the resources available to search among the myriad of possible combinations. Phenotypic performance of the array of possible combinations forms what can be viewed as an adaptation or fitness landscape (Cooper and Hammer 1996). Crop improvement then becomes a search strategy on that complex G × M × E landscape. However, currently we have difficulty reliably predicting (and navigating to) the desired destination on the adaptation landscape. We require prediction of phenotype based on genotype to underpin yield advance. In plant breeding, traditional empirical methods have involved measuring phenotypic performance of large segregating populations in multi-environment trials and applying rigorous statistical procedures based on quantitative genetic theory to identify superior individuals. This traditional phenotypic selection approach has been, and remains, successful for a number of crops, but cost per unit yield gain has risen substantially, interactions with management are not integrated, and genotype-by-environment interactions confound selection.

With recent progress in molecular technologies for genome sequencing and functional genomics, it had been widely expected that a gene-by-gene engineering approach would enable enhanced efficiency in crop improvement. Indeed, there have been successes in developing plants that better resist pests or tolerate herbicides. Those cases involved single-gene transformations where plant phenotypic response scaled directly from the level of molecular action. This could be described as a short 'phenotypic distance' (Fig. 7.1). However, little of this promise has been realised for key complex traits where relationships among components and their genetic controls involve quantitative multi-gene interactions. Integrating gene effects across scales of biological organisation in such situations is not straightforward. Complexities associated with gene interactions, mediated via transcriptional and post-transcriptional regulation, or distributed control of fluxes in plant metabolic pathways are major impediments to scaling from gene network to phenotype, so that phenotypic prediction based on a gene-by-gene approach remains elusive (Hammer et al. 2006; Benfey and Mitchell-Olds 2008).

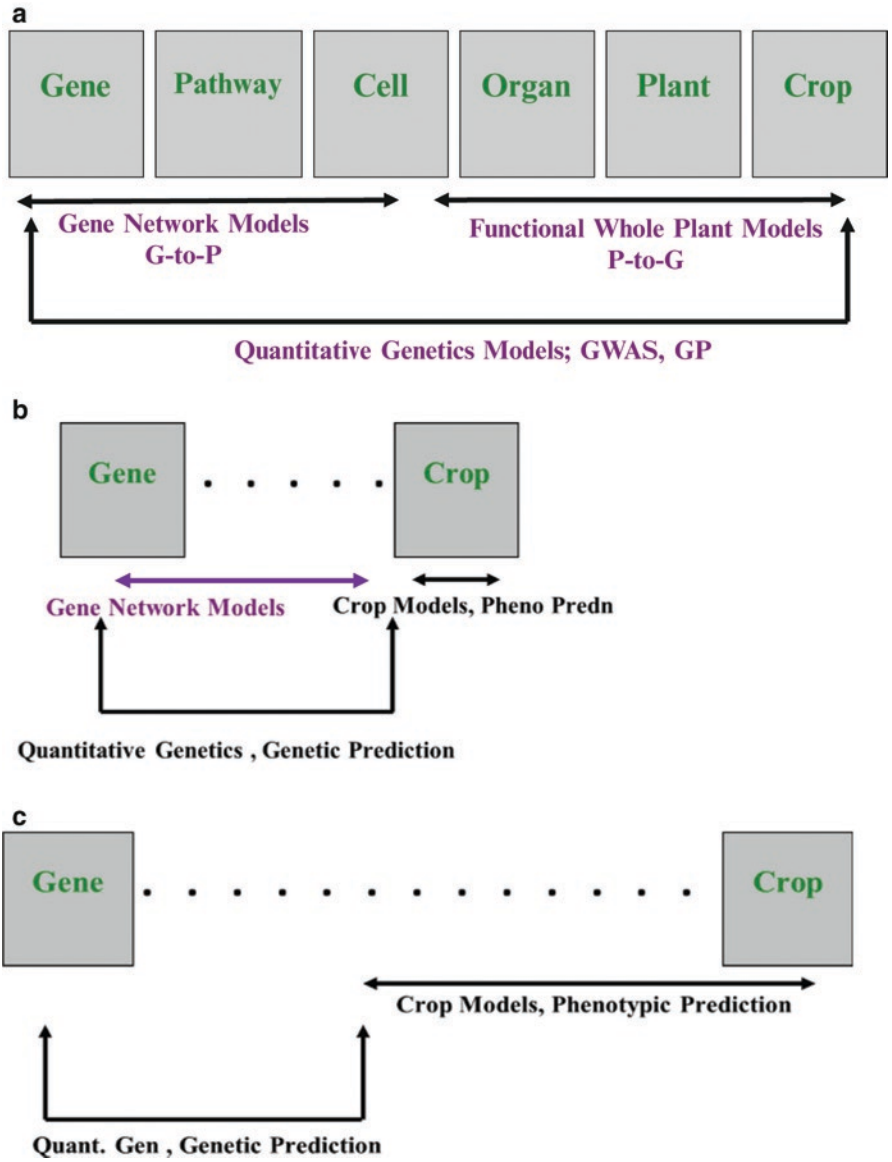


Fig. 7.1 (**a**) Approaches to G-to-P prediction, their association with levels of biological organisation, and the concepts of (b) 'short' and (c) 'long' phenotypic distance for traits that do, or do not, scale readily from molecular to whole organism scale. (**b**) Short phenotypic distance where traits scale directly from molecular to organism scale and there is a likely greater role for genetic prediction and gene network models. (**c**) Long phenotypic distance where traits do not scale readily from molecular to whole organism level and there is likely a greater role for ecophysiology and crop models

Developments in molecular genetic technologies have nonetheless allowed the focus of practical crop improvement to shift from the level of the individual (genotype) to the level of genomic region (e.g., quantitative trait locus – *QTL*) (Hammer and Jordan 2007). The ability to inexpensively and densely map genomes has facilitated development of molecular breeding strategies (Cooper et al. 2005, 2009). However, their applicability to complex traits remains constrained by context-dependent gene effects attributed to gene-gene and gene-environment interactions, which restrict predictive power of associations of genes/genomic regions with phenotypic responses. There is a long 'phenotypic distance' due to the extent of the biological integration required from the causal polymorphisms at genome scale to the phenotype of interest (e.g., Sinclair et al. 2004) (Fig. 7.1). Despite this limitation, Cooper et al. (2005) found that even though many of the context-dependent effects of genetic variation on phenotypic variation can reduce the rate of genetic progress from breeding, it is possible to design molecular breeding strategies for complex traits that on average will outperform phenotypic selection. Continuing advances in genotyping and crop genomics (Heffner et al. 2009; Morrell et al. 2012; Morris et al. 2012) have now facilitated association mapping approaches that assess correlation of phenotype with genotype in populations or panels of unrelated individuals. Such genome wide association studies rely on advanced statistical procedures to identify associations between a phenotype and a genomic marker profile. Genomic selection involves the use of phenotypic prediction equations based on profiles of marker data from a training set of genotypes, which have been phenotyped. The predictions are then applied across breeding materials that are genotyped extensively but not phenotyped. This offers considerable potential for more rapid genetic gain in breeding. However, for complex traits, the procedure still suffers from context-dependent effects and the 'phenotypic distance' problem (Fig. 7.1). Association mapping and genomic selection rely on the stability of the relationship between a phenotype and the set of genomic markers found in the training set, which is strongly dependent on the relevance of the genotypes and environments sampled.

Here we consider concepts associated with genotype-to-phenotype (G-to-P) modelling and how whole plant/crop ecophysiology and functional whole plant modelling can provide an effective link between molecular and organism scales to enhance efficiency of molecular breeding and crop improvement. There are two main avenues highlighted. Firstly, we describe how to enhance phenotyping strategies by using ecophysiological insight derived from dynamic crop growth and development modelling. This involves dissecting complex traits to more robust targets by reducing 'phenotypic distance' and context dependencies. Secondly, we show how to use crop growth and development models for trait evaluation and phenotypic prediction. This requires robust dynamic crop growth and development models that can predict consequences of context-dependent genotype and environment effects in target production regions.

7.2 Genotype-to-Phenotype (G-to-P) Modelling

There is a range of approaches for G-to-P modelling for complex traits that can be somewhat simplistically represented in relation to broad levels of biological organisation (Fig. 7.1). Gene network models that account for gene expression dynamics and metabolic pathway interactions have potential to account for gene context dependencies but require advanced knowledge of network structure and dynamics (see Chap. 1 of this book by Baldazzi et al.). Model species (e.g., *Arabidopsis*) provide opportunities to capture such knowledge. However, the issue of scaling from network to whole plant phenotypic response remains, unless direct associations exist, as for example with transition to flowering where the network is well characterised and scaling is direct (van Oosterom et al. 2006; Salazar et al. 2009; Dong et al. 2012). Network models involving enzyme kinetics have also been developed for exploring aspects of starch synthesis as a means to focus efforts aimed at manipulating starch structure and functionality (Wu et al. 2013).

Functional whole plant models have potential to account for environment context dependencies as they attempt to encapsulate dynamic plant-environment interactions based on physiological understanding (Tardieu 2003; Reymond et al. 2003; Chenu et al. 2008; Yin and Struik 2008; Hammer et al. 2005, 2010). It is plausible to link the vector of coefficients defining the plant characteristics to genomic regions, but the issue of scaling from coefficient to gene level remains problematic. There are some examples where the 'metaprocesses' of ecophysiology, such as the ubiquitous canopy radiation use efficiency (RUE) (Sinclair and Muchow 1998), have been dissected to their physiological or metabolic underpinning processes, firstly via canopy photosynthesis models that are driven by photosynthesis-light response curves and canopy structure (Hammer and Wright 1994; de Pury and Farquhar 1997), and more recently by direct linkage of those models to biochemical pathway models for photosynthesis (Gu et al. 2014). Hence, as knowledge advances, there are opportunities for gene network and metabolic pathway models to interface with crop ecophysiological models and advance dynamic modelling capability to account for genetic and environmental context dependencies in G-to-P prediction.

7.3 Whole Plant Ecophysiology and Modelling

Plant/crop models have been used extensively to facilitate decision making by crop managers, and to aid in education, but Hammer et al. (2002) suggested that greater explanatory power was required for their effective application in understanding and advancing the genetic regulation of plant performance and plant improvement. This is now even more prescient in the genomics era. Agronomic models contain a mix of descriptive and explanatory approaches that suffice for their application in decision/discussion support for crop management. Adequate prediction of resource use, crop growth and yield can be obtained with algorithms that describe aspects of crop

growth, such as plant leaf area, as a function of thermal time or plant leaf size distribution. The coefficients of these algorithms can be mapped to genomic regions, but this is unlikely to diminish any context dependencies, i.e., the coefficients will retain the context dependencies of the phenotypic variable they describe.

A physiological framework that facilitates further dissection and modelling of traits provides an avenue to overcome this problem. By enhancing the explanatory power of the modelling approach while not introducing undue complexity, it is possible to have phenotypic attributes as emergent properties of the model dynamics. This approach holds considerable promise for effective linking of genotype to phenotype and hence, molecular biology/genetics with crop improvement. Recent developments within the APSIM modelling platform (Hammer et al. 2010) have focused on structuring a generic cereal template to better accommodate the hierarchy of physiological determinants of crop growth and development needed for this more explanatory approach to plant modelling. They detail a case of the stay-green phenotype in sorghum (i.e., extended retention of green leaf area during grain filling), which was generated as an emergent consequence of canopy nitrogen (N) dynamics associated with genetic differences in dwarfing. Taller genotypes grew more and required more N for structural stem tissue, leaving less available for leaves, which was more rapidly diminished by translocation to grain during grain-filling (Fig. 7.2). Hence, stay-green was generated as an emergent consequence in the shorter genotypes in response to genetic differences in plant height.

Robust explanatory plant models have the potential to underpin G-to-P prediction by linking their coefficients with the genomic regions known to be associated with complex traits. However, to be effective, the linkage to model coefficients must reduce (or remove) the environmental and genetic context dependencies related to the phenotypic trait(s) that they generate. For example, the seasonal pattern of leaf area development is critical to resource (e.g., light, water) capture, and hence to crop growth and timing of stress. Studies at organ level (Reymond et al. 2003; Tardieu et al. 2005) on leaf expansion rate (LER) in maize have found that stable QTLs could be identified for responses of LER to temperature, vapour pressure deficit and plant water status, whereas QTLs for leaf area were dependent on the growing environment. Hence, by moving to the level of LER, environment context dependencies were removed. Some of the genomic regions associated with LER were also associated with silk extension and grain set in maize (Welcker et al. 2007). By enhancing the APSIM cereal template to operate at this level and incorporate genomic associations on LER and grain set, Chenu et al. (2009) were able to quantify impact at the crop yield level of the QTLs involved for a range of drought and climate scenarios.

7.4 Enhancing Breeding Efficiency

As indicated earlier, there are two main avenues by which crop ecophysiology and modelling can enhance breeding efficiency. The first involves use of ecophysiological insight from dynamic models to enhance phenotyping strategies by dissecting

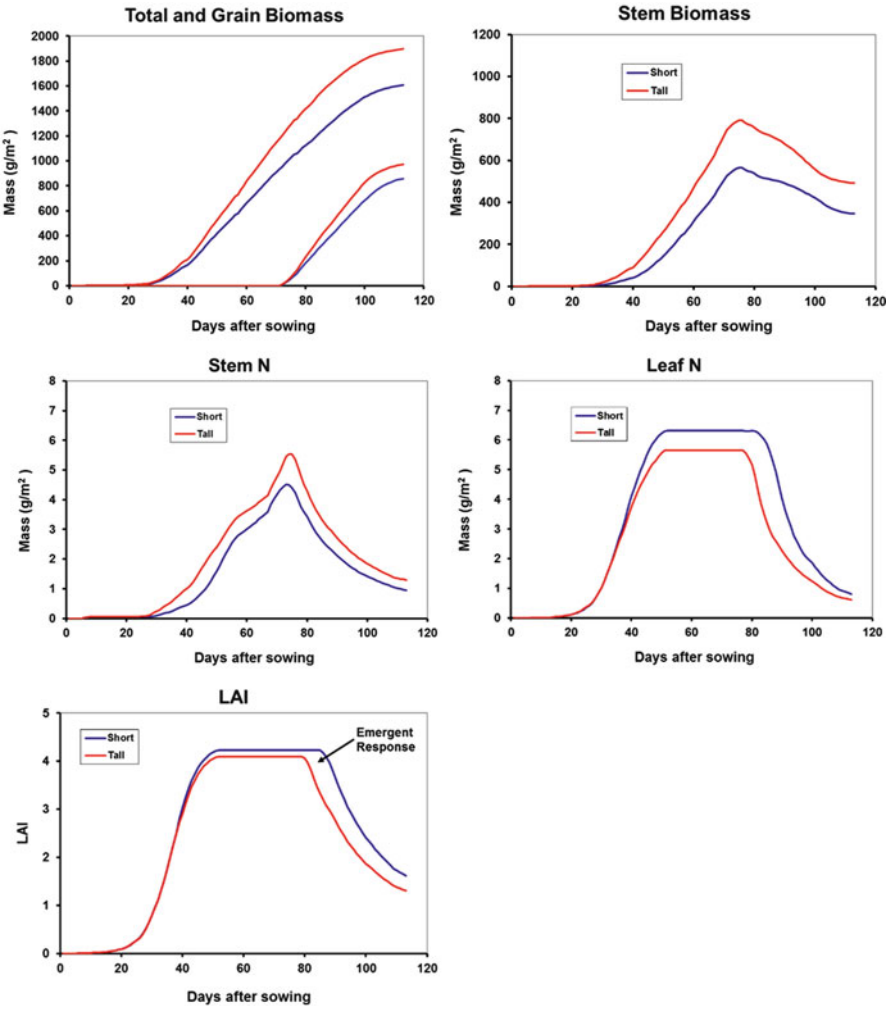


Fig. 7.2 Simulation of stay-green phenotype in sorghum as an emergent consequence of nitrogen (N) dynamics associated with differences in dwarfing genes via effects on structural N requirement for stem. Panels show organ biomass, organ N, and canopy leaf area index (LAI) simulated throughout the crop cycle for hybrids differing in height. The emergent delayed onset of senescence (i.e., 'stay-green') of the short hybrid is indicated on the panel giving the dynamics of canopy LAI. After Hammer et al. (2010)

complex traits to more robust targets that help to deal with phenotypic distance. The second involves using crop growth and development models for trait evaluation and phenotypic prediction in target production regions to help prioritise effort and assess breeding strategies. We consider an example of each.

7.4.1 *Phenotyping for Drought Adaptation: Water Capture by Root Systems*

The dynamics of water capture by root systems through the crop life cycle is critical to drought adaptation in water-limited environments. Slight changes in availability of soil moisture reserves associated with root system architecture, and in the timing of that availability, can have major consequences on yield in terminal drought environments, as suggested in wheat (Manschadi et al. 2006: Kirkegaard et al. 2007). In studies on sorghum in large rhizotrons, nodal root angle in young sorghum plants was shown to influence vertical and horizontal root distribution of mature plants in the soil profile and, hence, their ability to extract soil water (Singh et al. 2012). Types with narrower root angle tended to explore the soil profile more effectively at depth. These results suggested that genetic variation in nodal root angle of young sorghum plants could be a useful selection criterion for specific drought adaptation. Singh et al. (2010) had discovered this variation in nodal root angle when conducting studies on the morphological and architectural development of sorghum root systems in a small number of genotypes. They noted that due to the relatively late timing of appearance of nodal roots in sorghum, screening for genetic variation in the trait would require a small chamber system to grow plants until at least six leaves had fully expanded. They subsequently developed and implemented such a phenotyping system (Fig. 7.3) (Singh et al. 2011) and identified significant genotypic variation in the flush angle of nodal roots for a diverse set of sorghum genotypes.

Pursuing genetic variation in this trait by phenotyping mapping populations using this system, Mace et al. (2012) identified four QTLs for nodal root angle in sorghum that explained 58.2 % of the phenotypic variance and were validated across a range of diverse inbred lines. Three of the four nodal root angle QTLs showed homology to previously identified root angle QTLs in rice and maize, whereas all four QTLs co-located with previously identified QTLs for the drought adaptation trait stay-green in sorghum. Simulation studies based on possible differences in root architecture and their estimated effect on extent of water capture by root systems suggested significant yield advantage (up to 15 %) in low-yielding situations in a key sorghum production environment in NE Australia (Fig. 7.4). A putative association between nodal root angle QTLs and grain yield, which was consistent with the simulation studies, was identified through single marker analysis on field testing data from a subset of the mapping population grown in hybrid combination with three different tester lines. The identification of nodal root angle QTLs presents new opportunities for improving drought adaptation mechanisms via molecular breeding to manipulate a trait for which selection has previously been very difficult.

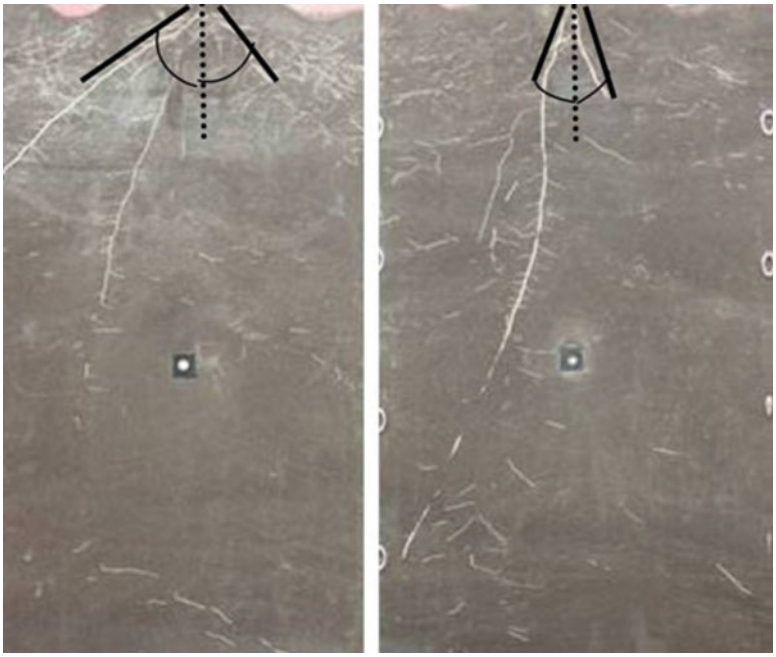


Fig. 7.3 Sorghum root angle phenotyping system. Individual plants were grown in specially designed root observation chambers until the sixth leaf had fully expanded. Each chamber consisted of two 50 cm high and 45 cm wide perspex sheets with the 3 mm gap filled with a dark, fine sandy soil so roots were clearly visible. Two genotypes divergent in angle of the first flush of nodal roots are shown. For mapping studies, chambers were stacked into tubs, covered below the plant to exclude light, and watered with a complete hydroponic solution daily. After Singh et al. (2011)

7.4.2 Phenotypic Prediction: Evaluating G × M × E Effects on Drought Adaptation

The timing of water availability through the crop life cycle is critical to drought adaptation in water-limited environments. The key principle involved is maximising the amount of water captured by the plant as transpiration (i.e., productive water use), while optimising its distribution pre- and post-anthesis. Appropriate plant trait and management system combinations for specific situations can be designed using this principle (Hammer 2006; Hammer et al. 2014). Reduced plant population and skip-row planting systems are common agronomic practices implemented with the intent of avoiding water limitation at anthesis and increasing the proportion of water use during the reproductive phase (Lyon et al. 2003; Whish et al. 2005). Genotypes expressing reduced early growth (Ray et al. 1997), early maturity (Ravi Kumar et al. 2009), or reduced tillering (van Oosterom et al. 2011; Alam et al. 2014a, b; Borrell et al. 2014a, b) can all provide a path towards soil water conservation and yield increase under drought stress (also see Chap. 5 of this book by Luquet et al.).

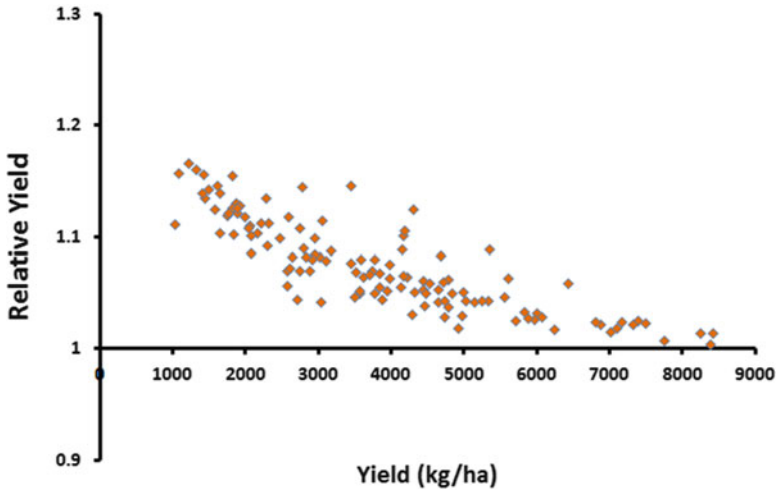


Fig. 7.4 Simulated yield of narrow root angle sorghum genotype relative to simulated yield of a standard type versus the yield of the standard type (kg ha^{-1}) for a 100-year simulation at Goondiwindi, NE Australia, assuming conventional agronomy. Each point represents 1 year of the simulation. The narrow root angle type had the same rooting depth but was assumed to access up to 15 mm extra soil water below 1 m soil depth if it was available in the soil profile

Hammer et al. (2014) reported a study that simulated the complex phenotypic adaptation landscape for combinations of these G and M factors for sorghum in the mostly water-limited production environments of north-eastern Australia, where sorghum is commonly grown as a row crop. Attributes for M employed in the simulations included three types of row configuration (solid 1 m rows; single skip row; double skip row; Whish et al. 2005) and four levels of plant density (3.5, 5.0, 6.5, 8.0 plants m^{-2}) while G attributes included nine levels of maturity (Ravi Kumar et al. 2009; Mace et al. 2013), and nine levels of tillering (Kim et al. 2010; Alam et al. 2014a, b). Levels of maturity were introduced by varying the time to floral initiation within the range −30 to +30 °Cd relative to the standard hybrid (with value 160 °Cd) using steps of 7.5 °Cd to generate the nine types. In addition to the effect on crop duration, this generates a change in total leaf number and hence modifies the pattern of leaf area development through the crop life cycle (Hammer et al. 1993). The range employed generates difference from the standard hybrid (17 leaves) within the range −1.5 to +1.5 total leaf number (Ravi Kumar et al. 2009). Levels of tillering were introduced by adding to, or subtracting from, the fertile tiller numbers assigned to the standard hybrid, within the range −2 to +2 tillers using steps of 0.5 fertile tillers to generate the nine types. For the lowest tillering type, this generates a plant that is uniculm in nearly all situations. Fertile tiller number affects maximum potential plant leaf area and hence the pattern of leaf area development through the crop life cycle (Hammer et al. 1993).

Components of the simulated yield adaptation landscapes can be viewed using heat maps of yield across a number of variables. Figure 7.5 presents yield levels for

two consecutive years at one key location (Emerald) illustrating grain yield landscapes associated with variation in tillering, maturity, row configuration, and planting density. In 2005 (Fig. 7.5a), which was lower-yielding, the highest yield occurred with a low tillering, late maturing type, grown at low population in a single skip row configuration. In contrast, in 2006, with the same sowing date and soil conditions (Fig. 7.5b), yields were greater, with the maximum yield occurring with a high tillering, relatively early maturing type, grown at high population in a solid row configuration. This contrast demonstrates the instability in the adaptation landscape with different combinations of G×M being favoured depending on E, and highlights the difficulty in seeking broad adaptation in such variable production environments. Hammer et al. (2014) used the simulated phenotypic landscape to evaluate the extent of the potential advantage of a breeding strategy pursuing specific adaptation versus one pursuing broad adaptation across all environments. While significant advantages to specific adaptation were identified, this would introduce more cost and complexity to breeding.

Other recent simulation studies have also suggested that a limited maximum transpiration rate may contribute to early season water conservation, and as a consequence to improved yield under drought (Sinclair et al. 2005, 2010). The consequence of a maximum rate of transpiration is that around midday, when vapour pressure deficit (VPD) is high, plants would not lose water at an unrestricted rate (see Chap. 6 of this book by Sinclair et al.). This limitation would be manifested in decreased stomatal conductance during periods of high VPD. This behaviour would generate increased transpiration efficiency (TE, biomass accumulated per unit water transpired) for the crop because of decreased gas exchange during periods of high demand for crop water use. Experimental studies have identified a limited maximum transpiration trait in a range of species by quantifying responses to VPD (Fletcher et al. 2007; Sadok and Sinclair 2009a, b; Kholova et al. 2010; Jyostna Devi et al. 2010; Gholipoor et al. 2010, 2013; Yang et al. 2012; Choudhary and Sinclair 2014; Choudhary et al. 2014).

7.5 Implications

These examples demonstrate two of the main ways that whole plant ecophysiology and modelling can enhance molecular breeding via improved G-to-P understanding and prediction:

- Physiological dissection of complex traits in a dynamic framework – Experimental studies in controlled genetic backgrounds provide the means to determine and quantify the functional biology underpinning phenotypic differences, and thus inform high throughput phenotyping. Dynamic process concepts in crop models provide the analytical context to frame that understanding. Attributes can then be linked to genomic regions (QTLs) in a way that reduces context dependency and phenotypic distance and generates coefficients for dynamic crop models that quantify ecophysiological implications of genetic regulation.

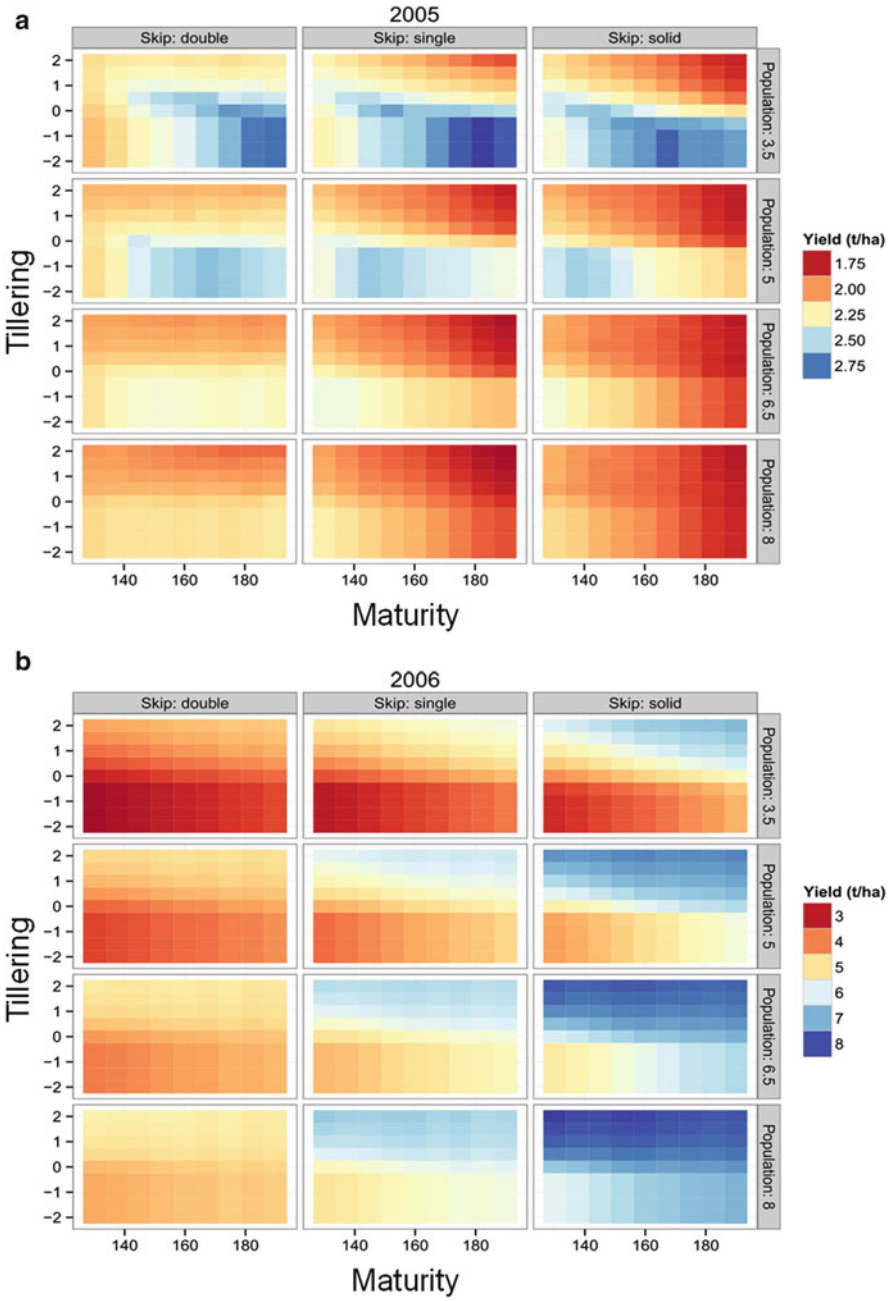


Fig. 7.5 Simulated phenotypic landscapes of sorghum grain yield (t/ha) at Emerald (NE Australia) in (**a**) 2005 and (**b**) 2006 for genotypes varying in tillering (positive values for high tillering types) and maturity (high thermal time requirement values for late maturing types – see text) and crop management varying in row configuration (double skip, single skip, and solid 1 m rows) and density (3.5–8 plants m^{-2}). After Hammer et al. (2014)

- Predicting consequences of genetic variation – Crop models with trait physiology and/or genetics embedded in their coefficient structure can be implemented in a predictive context to estimate by simulation the likely relevance of genetic variation for specific environments and management systems (i.e., G×M×E). This simulated phenotypic value has the potential to provide a basis for estimating trait value and weighting genomic regions in molecular breeding in a manner that is more robust than empirical genomic prediction approaches.

In both of these example cases, incorporating explanatory sub-models based on physiological insight into the quantitative crop model provided a basis to link changes at genomic regions directly to their emergent phenotypic consequences at the crop level via intermediary traits in a way that reduced context dependencies and phenotypic distance. Such an approach provides a pathway to effective applications in molecular breeding (Cooper et al. 2014a, b). Further, the functional whole plant models can be used to explore breeding strategies by generating the adaptation landscape of possible G×M×E combinations on which breeding system simulation tools can map the trajectories resulting from specific breeding approaches (Cooper et al. 2002; Chapman et al. 2003; Hammer et al. 2005; Messina et al. 2009, 2011). In this way, whole plant physiology and modelling can provide an effective link between molecular knowledge, genotyping capacity, and the practice of crop improvement.

Acknowledgements This paper summarises the research of a team of people, which would not have been possible without their enthusiasm and dedication nor without the financial support of a number of funders including Australian Research Council, Grains Research and Development Corporation, DuPont Pioneer, European Union FP7, and Australian Centre for International Agricultural Research. The authors acknowledge Crop and Pasture Science (CSIRO Publishing) and Journal of Experimental Botany (Oxford Journals) for permissions to re-use figures.

References

Alam MM, Hammer GL, van Oosterom EJ, Cruickshank A, Hunt C, Jordan DR (2014a) A physiological framework to explain genetic and environmental regulation of tillering in sorghum. New Phytol 203:155–167

Alam MM, Mace ES, van Oosterom EJ, Cruickshank A, Hunt CH, Hammer GL, Jordan DR (2014b) QTL analysis in multiple sorghum populations facilitates the dissection of the genetic and physiological control of tillering. Theor Appl Genet 127:2253–2266

Benfey PN, Mitchell-Olds T (2008) From genotype to phenotype: systems biology meets natural variation. Science 320:495–497

Borrell AK, Mullet JE, George-Jaeggli B, van Oosterom EJ, Hammer GL, Klein PE, Jordan DR (2014a) Drought adaptation of stay-green in sorghum associated with canopy development, leaf anatomy, root growth and water uptake. J Exp Bot 65:6251–6263

Borrell AK, van Oosterom EJ, Mullet JE, George-Jaeggli B, Jordan DR, Klein PE, Hammer GL (2014b) Stay-green alleles individually enhance grain yield in sorghum under drought by modifying canopy development and water uptake patterns. New Phytol 203:817–830

Chapman SC, Cooper M, Podlich D, Hammer GL (2003) Evaluating plant breeding strategies by simulating gene action and dryland environment effects. Agron J 95:99–113

Chenu K, Chapman SC, Hammer GL, McLean G, Ben Haj Salah H, Tardieu F (2008) Short-term responses of leaf growth rate to water deficit scale up to whole-plant and crop levels: an integrated modelling approach in maize. Plant Cell Environ 31:378–391

Chenu K, Chapman SC, Tardieu F, McLean G, Welcker C, Hammer GL (2009) Simulating the yield impacts of organ-level quantitative trait loci associated with drought response in maize – a "gene-to-phenotype" modeling approach. Genetics 183:1507–1523

Choudhary S, Sinclair TR (2014) Hydraulic conductance differences among sorghum genotypes to explain variation in restricted transpiration rates. Funct Plant Biol 41:270–275

Choudhary S, Sinclair TR, Messina CD, Cooper M (2014) Hydraulic conductance of maize hybrids differing in transpiration response to vapor pressure deficit. Crop Sci 54:1147–1152

Cooper M, Hammer GL (1996) Synthesis of strategies for crop improvement. In: Cooper M, Hammer GL (eds) Plant adaptation and crop improvement. CAB International, ICRISAT & IRRI, Wallingford, pp 591–623

Cooper M, Chapman SC, Podlich DW, Hammer GL (2002) The GP problem: quantifying gene-to-phenotype relationships. In Silico Biol 2:151–164. http://www.bioinfo.de/journals.html

Cooper M, Podlich DP, Smith OS (2005) Gene-to-phenotype models and complex trait genetics. Aust J Agr Res 56:895–918

Cooper M, van Eeuwijk FA, 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

Cooper M, Gho C, Leafgren R, Tang T, Messina C (2014a) Breeding drought-tolerant maize hybrids for the US corn-belt: discovery to product. J Exp Bot 65:6191–6204

Cooper M, Messina CD, Podlich D, Totir LR, Baumgarten A, Hausmann NJ, Wright D, Graham G (2014b) Predicting the future of plant breeding: complementing empirical evaluation with genetic prediction. Crop Pasture Sci 65:311–336

de Pury DGG, Farquhar GD (1997) Simple scaling of photosynthesis from leaves to canopies without the errors of big-leaf models. Plant Cell Environ 20:537–557

Dong Z, Danilevskaya O, Abadie T, Messina C, Coles N, Cooper M (2012) A gene regulatory network model for floral transition of the shoot apex in maize and its dynamic modeling. PLoS One 7:e43450

Fletcher AL, Sinclair TR, Allen LH Jr (2007) Transpiration responses to vapor pressure deficit in well watered and 'slow wilting' and commercial soybean. Environ Exp Bot 61:145–151

Gholipoor M, Prasad PVV, Mutava RN, Sinclair TR (2010) Genetic variability of transpiration response to vapor pressure deficit among sorghum genotypes. Field Crops Res 119:85–90

Gholipoor M, Choudhary S, Sinclair TR, Messina CD, Cooper M (2013) Transpiration response of maize hybrids to atmospheric vapour pressure deficit. J Agron Crop Sci 199:155–160

Gu J, Yin X, Stomph T-J, Struik PC (2014) Can exploiting natural genetic variation in leaf photosynthesis contribute to increasing rice productivity? A simulation analysis. Plant Cell Environ 37:22–34

Hammer G (2006) Pathways to prosperity: breaking the yield barrier in sorghum. Agric Sci 19:16–22

Hammer GL, Jordan DR (2007) An integrated systems approach to crop improvement. In: Spiertz JHJ, Struik PC, van Laar HH (eds) Scale and complexity in plant systems research: gene-plant-crop relations, vol 21, Wageningen UR – Frontis series. Springer, The Netherlands, pp 45–61

Hammer GL, Wright GC (1994) A theoretical analysis of nitrogen and radiation effects on radiation use efficiency in peanut. Aust J Agr Res 45:575–589

Hammer GL, Carberry PS, Muchow RC (1993) Modelling genotypic and environmental control of leaf area dynamics in grain sorghum. I. Whole plant level. Field Crops Res 33:293–310

Hammer GL, Kropff MJ, Sinclair TR, Porter JR (2002) Future contributions of crop modelling – from heuristics and supporting decision-making to understanding genetic regulation and aiding crop improvement. Eur J Agron 18:15–31

Hammer G, Chapman S, van Oosterom E, Podlich D (2005) Trait physiology and crop modelling as a framework to link phenotypic complexity to underlying genetic systems. Aust J Agr Res 56:947–960

Hammer G, Cooper M, Tardieu F, Welch S, Walsh B, van Eeuwijk F, Chapman S, Podlich D (2006) Models for navigating biological complexity in breeding improved crop plants. Trends Plant Sci 11:587–593

Hammer GL, van Oosterom E, McLean G, Chapman SC, Broad I, Harland P, Muchow RC (2010) Adapting APSIM to model the physiology and genetics of complex adaptive traits in field crops. J Exp Bot 61:2185–2202

Hammer GL, McLean G, Chapman S, Zheng B, Doherty A, Harrison MT, van Oosterom E, Jordan D (2014) Crop design for specific adaptation in variable dryland production environments. Crop Pasture Sci 65:614–626. http://www.publish.csiro.au/nid/40/paper/CP14088.htm

Heffner EL, Sorrells ME, Jannink JL (2009) Genomic selection for crop improvement. Crop Sci 49:1–12

Jyostna Devi M, Sinclair TR, Vadez V (2010) Genotypic variation in peanut for transpiration response to vapor pressure deficit. Crop Sci 50:191–196

Kholova J, Hash CT, Kumar PL, Yadav RS, Kočová M, Vadez V (2010) Terminal drought-tolerant pearl millet [*Pennisetum glaucum* (L.) R. Br.] have high leaf ABA and limit transpiration at high vapour pressure deficit. J Exp Bot 61:1431–1440

Kim HK, Luquet D, van Oosterom E, Dingkuhn M, Hammer G (2010) Regulation of tillering in sorghum: genotypic effects. Ann Bot 106:69–78

Kirkegaard JA, Lilley JM, Howe GN, Graham JM (2007) Impact of subsoil water use on wheat yield. Aust J Agr Res 58:303–315

Lyon DJ, Hammer GL, McLean GB, Blumenthal JM (2003) Simulation supplements field studies to determine no-till dryland corn population recommendations for semiarid western Nebraska. Agron J 95:884–891

Mace ES, Singh V, van Oosterom EJ, Hammer GL, Hunt CH, Jordan DR (2012) QTL for nodal root angle in sorghum (*Sorghum bicolor* L. Moench) co-locate with QTL for traits associated with drought adaptation. Theor Appl Genet 124:97–109

Mace ES, Hunt CH, Jordan DR (2013) Supermodels: sorghum and maize provide mutual insight into the genetics of flowering time. Theor Appl Genet 126:1377–1395

Manschadi AM, Christopher J, deVoil P, Hammer GL (2006) The role of root architectural traits in adaptation of wheat to water-limited environments. Funct Plant Biol 33:823–837

Messina C, Hammer G, Dong Z, Podlich D, Cooper M (2009) Modelling crop improvement in a G*E*M framework via gene-trait-phenotype relationships. In: Sadras VO, Calderini D (eds) Crop physiology: applications for genetic improvement and agronomy. Academic/Elsevier, The Netherlands, pp 235–265

Messina CD, Podlich D, Dong Z, Samples M, Cooper M (2011) Yield–trait performance landscapes: from theory to application in breeding maize for drought tolerance. J Exp Bot 62:855–868

Morrell PL, Buckler ES, Ross-Ibarra J (2012) Crop genomics: advances and applications. Nat Rev Genet 13:85–96

Morris GP, Ramu P, Deshpande SP et al (2012) Population genomic and genome-wide association studies of agroclimatic traits in sorghum. Proc Natl Acad Sci U S A 110:453–458

Ravi Kumar S, Hammer GL, Broad I, Harland P, McLean G (2009) Modelling environmental effects on phenology and canopy development of diverse sorghum genotypes. Field Crops Res 111:157–165

Ray JD, Samson BK, Sinclair TR (1997) Vegetative growth and soil water extraction of two maize hybrids during water deficits. Field Crops Res 52:135–142

Reymond M, Muller B, Leonardi A, Charcosset A, Tardieu F (2003) Combining quantitative trait loci analysis and an ecophysiological model to analyze the genetic variability of the responses of maize leaf growth to temperature and water deficit. Plant Physiol 131:664–675

Sadok W, Sinclair TR (2009a) Genetic variability of transpiration response to vapor pressure deficit among soybean (*Glycine max* [L.] Merr.) genotypes selected from a recombinant inbred line population. Field Crops Res 113:156–160

Sadok W, Sinclair TR (2009b) Genetic variability of transpiration response to vapor pressure deficit among soybean cultivars. Crop Sci 49:955–960
Salazar JD, Saithong T, Brown PE et al (2009) Prediction of photoperiodic regulators from quantitative gene circuit models. Cell 139:1170–1179
Sinclair TR, Muchow RC (1998) Radiation use efficiency. Adv Agron 65:215–265
Sinclair T, Purcell L, Sneller CH (2004) Crop transformation and the challenge to increase yield potential. Trends Plant Sci 9:70–75
Sinclair TR, Hammer GL, van Oosterom EJ (2005) Potential yield and water-use efficiency benefits in sorghum from limited maximum transpiration rate. Funct Plant Biol 32:945–952
Sinclair TR, Messina CD, Beatty A, Samples M (2010) Assessment across the United States of the benefits of altered soybean drought traits. Agron J 102:475–482
Singh V, van Oosterom EJ, Jordan DR, Hammer GL (2010) Morphological and architectural development of root systems in sorghum and maize. Plant and Soil 333:287–299
Singh V, van Oosterom EJ, Jordan DR, Hammer GL (2011) Genetic variability and control of root architecture in sorghum. Crop Sci 51:2011–2020
Singh V, van Oosterom EJ, Jordan DR, Hammer GL (2012) Genetic control of nodal root angle in sorghum and its implications on water extraction. Eur J Agron 42:3–10
Tardieu F (2003) Virtual plants: modelling as a tool for the genomics of tolerance to water deficit. Trends Plant Sci 8:9–14
Tardieu F, Reymond M, Muller B, Simonneau T, Sadok W, Welcker C (2005) Linking physiological and genetic analyses of the control of leaf growth under changing environmental conditions. Aust J Agr Res 56:937–946
van Oosterom EJ, Hammer GL, Chapman SC, Doherty A, Mace E, Jordan DR (2006) Predicting flowering time in sorghum using a simple gene network: functional physiology or fictional functionality? In: Borrell AK et al. (eds) Proceedings of the 5th Australian sorghum conference, Gold Coast, 30 Jan–2 Feb 2006
van Oosterom EJ, Borrell AK, Deifel K, Hammer GL (2011) Does increased leaf appearance rate enhance adaptation to post-anthesis drought stress in sorghum? Crop Sci 51:2728–2740
Welcker CB, Boussuge CB, Ribaut JM, Tardieu F (2007) Are source and sink strengths genetically linked in maize plants subjected to water deficit? A QTL study of the responses of leaf growth and of anthesis-silking interval to water deficit. J Exp Bot 58:339–349
Whish J, Butler G, Castor M, Cawthray S, Broad I, Carberry P, Hammer G, McLean G, Routley R, Yeates S (2005) Modelling the effects of row configuration on sorghum yield in north-eastern Australia. Aust J Agr Res 56:11–23
Wu AC, Morell MK, Gilbert RG (2013) A parameterized model of amylopectin synthesis provides key insights into the synthesis of granular starch. PLoS One 8(6):e65768
Yang Z, Sinclair TR, Zhu M, Messina CD, Cooper M, Hammer GL (2012) Temperature effect on transpiration response of maize plants to vapour pressure deficit. Environ Exp Bot 78:157–162
Yin X, Struik PC (2008) Applying modelling experiences from the past to shape crop systems biology: the need to converge crop physiology and functional genomics. New Phytol 179:629–642

Chapter 8
Crop Modeling Approaches for Predicting Phenotype of Grain Legumes with Linkage to Genetic Information

Kenneth J. Boote, C. Eduardo Vallejos, James W. Jones, and Melanie J. Correll

Abstract In this chapter we introduce concepts on how mechanistic crop simulation models can be linked with genetic information to predict phenotype in different environments. There has been rapid advancement of genotyping along with continued improvement of mechanistic simulation models that predict dynamic daily (or faster) growth processes of crops in response to varying weather, soils, and management conditions. Crop models have genotype-specific-parameters (GSPs) that describe performance of different cultivars; nevertheless, those GSPs are empirical mathematical parameters that are estimated directly from field phenotyping data. There is great opportunity to link the modeled GSPs with genes (or QTLs) obtained from genotyping, so that phenotypic performance can be directly predicted from genotype. The largest challenge is the phenotyping needed to characterize the phenotypes that result from gene expression in different environments. Examples are given of phenotyping in a recombinant common bean study. Additional mechanisms and different GSPs may be needed in the crop models to achieve this goal. Since crop models are already programmed to account for weather, soils, and management effects, they are efficient tools in which hypothetical alleles of genes can be evaluated for multiple environments. Model simulations illustrate examples of genotype-by-environment (G×E) interactions, where a given allele (gene) for a trait may have either positive or negative effects on yield, depending on weather and management conditions. Examples of linkage of GSPs to genes are given for common bean, along with phenotypic outcomes of growth patterns observed to be very responsive to presence or absence of alleles of genes.

K.J. Boote (✉)
Agronomy Department, University of Florida,
P.O. Box 110500, 32611-0500 Gainesville, FL, USA
e-mail: kjboote@ufl.edu

C.E. Vallejos
Horticultural Sciences Department, University of Florida, 32611 Gainesville, FL, USA

J.W. Jones • M.J. Correll
Agricultural and Biological Engineering Department, University of Florida,
32611 Gainesville, FL, USA

X. Yin, P.C. Struik (eds.), *Crop Systems Biology*,
DOI 10.1007/978-3-319-20562-5_8

8.1 Introduction

The processes at the core of mechanistic crop simulation models are generic at the level of describing photosynthesis, respiration, translocation, and partitioning (Boote et al. 1998). However, the attempt to mimic the growth and development of a given crop species or a particular genotype using these mechanistic models soon reveals the necessity of having genetic-specific information in them (Boote et al. 2001). Models that include this information have the ability to establish the environmental dependencies (temperature, photoperiod, water availability, etc.) of the various genetically-determined processes that shape the phenotype of the crop. The genetic information is what allows the model to simulate the unique growth and developmental patterns of each genotype under specific environments. Present crop models have a basic set of parameters and relationships that make the model specific to a given crop, and they additionally have genotype-specific parameters (GSPs) ranging from few to dozens that allow simulation of cultivar variation. With the continued advancements of tools in molecular genetics, a logical advancement in the use of crop simulation models is to predict or mimic the phenotype of a cultivar based on its genotype through next generation sequencing at the gene level or through quantitative trait loci (QTLs), which can be identified directly, or through linked genetic markers.

In this paper, we review the literature, address specific questions, and give examples of approaches for linking genotype to phenotype with mechanistic crop simulation models. The questions to be addressed include:

1. Can increasing levels of genetic detail and variability be incorporated into the present crop simulation models to better represent the dynamic interactions among all the components of the biological system, in response to different environmental factors?
2. How can the new genetic analysis tools help to link the genotype to the dynamic phenotype under different environments?
3. Are there approaches that can integrate the power of genetics with the dynamic simulation of crops provided by computer simulation models?

8.2 Are the Crop Models Ready?

To begin addressing the first question, we should consider whether a model takes into account all the components of the system it is attempting to simulate, and the dynamic interactions of these components. More specifically, does the model have sufficient mechanistic detail to simulate the expression of a genetic trait; this level of detail and mechanism may be lacking in relatively simple models, as models need to be more than just a summary description of processes, but rather the phenotype should be an emergent outcome of the genetic variation in processes as affected by environment (Parent and Tardieu 2014). For instance, genetic information has

been included in modeling through the simulation of genetic marker effects on leaf area expansion (Reymond et al. 2004; Chenu et al. 2009). Similar efforts are the inclusion, in models, of genes known to regulate the cultivar life cycle, such as those that control flowering, determinacy, and maturity (White and Hoogenboom 1996; Hoogenboom et al. 1997; Hoogenboom and White 2003; Hoogenboom et al. 2004; Messina et al. 2006). Along these lines, the most advanced gene-based model is the *Arabidopsis thaliana* photothermal flowering model, which takes into account the pattern of expression of several flowering genes as they respond to the changing photothermal environment (Wilczek et al. 2009). These examples clearly illustrate the need for incorporating genetic information into crop models in the future. The foci on only leaf area expansion and life cycle progress, while a good starting point, are only a small part of the whole integrated simulation of growth and yield. We concur with Yin and Struik (2010) that present simulation models will need to include additional mechanisms to address specific genetic traits and responses with more accurate outcomes. Hammer et al. (2010) proposed that enhancing models for their ability to predict phenotype from genotype would require algorithms that represent the underlying processes controlled by the genotype, and that simulation of model dynamics under varying environments should lead to emergent properties of the phenotypes (see examples discussed in Chap. 7 of this book by Hammer et al.). Likewise, Chenu et al. (2009) described genetic traits (for models) as those that are 'environmentally-independent'. Finally, the modular/sub-modular structure of present day crop simulation models gives them the flexibility to link well with genetic markers and to incorporate, at any given time, new components without altering the core of the program code (Jones et al. 2001, 2003).

8.3 Is Genetics Advanced Enough? Relationship of Phenotype to Genotype

8.3.1 From Genes to DNA

Gregor Mendel (1866) first used the phenotype to develop the concept of the genotype, and provided the conceptual framework to connect these properties of all living organisms. However, it was Johannsen (1911) who coined these terms and provided a full description of their meaning. The advent of molecular biology with the associated development of DNA technology has made it possible to realize the genotype concept into a DNA sequence. For instance, most of the genes used by Mendel to deduce laws of inheritance have now been characterized at the molecular level (Lester et al. 1997; Martin et al. 1997). These developments give us the opportunity to identify the genotype of an organism, via a DNA detection method, and use this information to deduce its phenotype. For instance, Mendel's dwarf pea phenotype is caused by mutation that involves a single nucleotide polymorphism (SNP) in a gene that synthesizes an active form of gibberellic acid, a growth

hormone (Martin et al. 1997). Thus, current DNA technology makes it possible to test for that particular SNP and allows us to use the genotype to predict the phenotype of a pea seedling. However, most phenotypes of interest to agricultural production are very complex because they are governed by many genes, each leading to unique spatio-temporal patterns of expression and environmental responses (Benfey and Mitchell-Olds 2008).

8.3.2 Genes and QTLs

Genes contribute to a quantitative trait which can then be defined as a quantitative trait locus (QTL). Karl Sax (1923) was the first to report linkage between a gene controlling seed color, a true Mendelian trait, and a gene affecting seed size in the common bean. Since that time, the development of molecular markers has facilitated the identification and mapping of QTLs (Lander and Botstein 1989). This approach is based on the construction of molecular marker linkage maps. These maps were originally constructed with DNA restriction fragment length polymorphism, a tedious and cumbersome technology (Schlötterer 2004), but the technology has evolved allowing the construction of relatively dense linkage maps with minimal lab work (Elshire et al. 2011). Although it might be considered overkill, some groups have found it practical to sequence the genomes of an entire segregating progeny (Xu et al. 2013). The basic approach of QTL mapping is to query the entire genome to determine whether there are significant associations between markers mapped to a specific location and the quantitative phenotypic trait of interest (plant height, leaf area, weight per seed, etc.). Very sophisticated statistical approaches have been devised for this endeavor (Lander and Botstein 1989; Zeng 1993, 1994; Kao et al. 1999). Of these approaches, multiple interval mapping (MIM) is perhaps one of the most powerful QTL analytical tools. This is because it can simultaneously fit multiple QTLs allowing for a better estimation of parameters, distinguishing tightly linked QTLs from QTLs with pleiotropic effects, detecting epistatic interactions (Zeng 1994; Kao et al. 1999). The QTL mapping approaches listed above are powerful tools for the analysis of static end-point phenotypes, but are limited in their ability to capture and translate to the dynamic aspects of phenotype development. These limitations, however, have been partially overcome by the development of functional mapping (FM), which integrates mathematical functions of growth and development into a mixed model FM that has the capability of detecting QTLs that control growth and developmental trajectories (Ma et al. 2002; Wu and Lin 2006; Li and Wu 2010). The inclusion of the time dimension in FM makes this QTL mapping approach somewhat analogous to the dynamic approach used by crop simulation models, although crop models attempt to link genetics directly to crop processes and their sensitivities to daily or hourly environment, rather than using a mathematical time-dependent function.

8.3.3 Analyzing G×E with Statistical Tools

One of the major foci of attention of plant breeding programs is the proper assessment of genotype-by-environment (G×E) interactions. Their present tools so far, have primarily been statistical models. Zhao et al. (2004) showed the use of FM to characterize environmental interactions of individual QTLs controlling developmental trajectories. Given the observation that different genotypes have different abilities to respond to the environment, Wang et al. (2013) have provided a statistical framework to connect QTLs to the plastic behavior exhibited by phenotypes in different environments. Malosetti et al. (2013) presented a comprehensive analysis of this topic encompassing a range from simple *fixed effect models to mixed-effect models where the genotype and error are considered as random effects*, and also the use of factorial regression as a means to describe and explain G×E interactions (also see Chap. 3 of this book by Bustos et al.). In summary, genetic analysis tools can provide the means to ex-post dissect complex traits, as well as the means to capture the effect of genes during growth and development, and the interactions of these genes with the surrounding environment. *But can they give a dynamic phenotype for a new environment?* We propose that crop models, with their built-in sensitivity to environment (weather, management, soils, etc.), with inclusion of genetic information, can predict dynamic phenotypes in new target environments, whereas the statistical models are limited to those prior environments included in the statistical analyses.

Despite the sophistication of the different statistical approaches used in the genetic analysis of complex traits and their interactions with the environment, they are limited in their ability to simulate the phenotypic plasticity of a genotype under the continuously changing environment experienced by a crop grown in the field. This is where linkage to process-oriented dynamic crop simulation models can be used to advantage. Advances in both crop simulation approaches and modern genetic analysis of complex traits strongly suggest the potential synergism of combining these two approaches to link the genotype to the phenotype. The challenge is to find a suitable procedure to incorporate genotypic information into crop simulation models, thus using the model's built-in responsiveness to environment, and leading to correct emergent outputs (dynamic phenotypes).

While the task of finding genes or the QTLs for individual traits has become easier and faster, the second challenge remains whether one can use these statistical methods to predict the phenotype resulting from given QTLs for plants grown in many different environments. Knowing that a gene results in a particular biochemical product or plant morphology does not mean that the integrated outcome from a specific phenotype is clearly understood. This is where connection to crop simulation models can be helpful, because given 'traits' such as more rapid rooting, faster leaf elongation, earlier flowering, less stomatal sensitivity to water stress, etc., can give different crop growth and yield outcomes depending on the field environment (G×E interaction). This challenge, e.g., the need to understand what causes G×E interaction, is addressed in this chapter. Another challenge to linking genes to crop models

is the lack of sufficient field phenotyping (White et al. 2012). Simple phenotyping of seedling plants in a chamber or glasshouse is inadequate to represent mature plants in stressful field environments. New technologies are being proposed for phenotyping (Furbank and Tester 2011), but we feel that connection of genes (QTLs) to performance in field environments remains the most important aspect for predicting future performance in new target field environments.

8.4 Present Crop Models Are Responsive to Weather, Soils, and Management Effects

Over the last 30 years, crop models have been coded and parameterized to respond to time-varying weather variables, soil characteristics, and crop cultural management, with a primary goal of using the models as tools for crop management in the field under time- and space-varying conditions. In this respect, crop models are relatively ready to account for environmental and management effects although continued model improvements are needed. An advantage of crop models is that they are dynamic and predict daily growth processes in response to daily environmental conditions of irradiance, temperature, daylength, water stress, nitrogen (N) supply, etc., and this can be used to understand G × E interactions when model traits are coupled to genetics. Crop life cycle progression is captured with sensitivity of rate of leaf node appearance and reproductive progression to temperature and daylength, which are parameterized by cardinal base and optimum temperatures, critical daylength, and daylength sensitivity. Cardinal temperature sensitivities of phenology are generally considered to be species traits (although some researchers have suggested cultivar differences in cardinal temperatures). In either case, the amount of thermal unit accumulation as well as critical daylength and daylength sensitivity are considered cultivar traits. Characterizing life cycle timing of flowering, onset of reproductive growth, duration and termination of reproductive growth are important for yield performance and fitting cultivars to a given growth environment. Photosynthetic response to solar irradiance is linked to light interception by the canopy foliage along with quantum efficiency and light-saturated leaf assimilation rate, which also have dependence on temperature, CO_2, leaf N, and irradiance history. Actual dry matter accumulation depends on growth conversion efficiencies and maintenance respiration coefficients, of which the latter are temperature-dependent. Rate of leaf node appearance, leaf area expansion, height increase, root depth progression, addition of reproductive sites, and growth of individual grains, each have temperature dependency curves characterized by a base temperature, optimum temperature, and upper failure temperature. The crop models also carry a type of internal crop calendar or 'controller' by which the modeled crop controls the shifts and priorities of assimilate partitioning among root, leaf, stem, and reproductive tissues. In addition, the plant N-balance is regulated by the target (critical) N concentrations for growth of different organs along with the rate of N mobilization to reproductive organs.

These control algorithms are dependent on temperature, water, and soil N availability. Genetic variation will certainly influence most of these crop processes. The reason for highlighting crop responsiveness to environment, soils, and management, is to illustrate that crop models, when coupled to genetic traits, are ready-made tools to understand G×E interactions without the need to pre-suppose direct G×E genes. We will later illustrate examples where certain genetic traits may be positive for crop yield in one environment but negative in another environment. In other words, the G×E interaction is an emergent property of a single gene action (or multiple ones). This is a re-affirmation of the need to dig deeper to find the underlying processes influencing the phenotypic outcomes.

8.5 Genes, Processes, and Emergent Outcomes in Crop Models

One of the important issues with genotyping to phenotyping, is the definition of the action of the gene and what is the corresponding 'phenotype'. We quote from Chenu et al. (2009) who stated that "Using such 'environmentally stable' QTLs to model leaf elongation rate avoids complex QTL-environment interactions that are commonly observed for directly measured traits such as leaf length." Yin and Struik (2010) also distinguished between simple and complex phenotypic outcomes in their modeling. For example, by their definition, an integrated outcome such as yield, nutrient use efficiency, height, or total biomass would clearly be a 'complex' phenotype. It should be obvious that the phenotypic outcomes, such as leaf length or leaf width or seed size or height or final yield, are integrated cumulative outcomes of many processes occurring in fluctuating environments over time, so one must consider the gene (QTL) effects to be at the process level, not at the outcome/product level. While there are alleles with direct visible outcomes (flower color, dwarfism), most genes affecting quantitative outcomes cannot be directly equated to outcome 'traits'. This is an important issue where the emergent quantitative phenotype is determined by many genes beyond the ones being studied and the 'gene actions' are influenced differentially by environment or give different benefits depending on environment. For example, Chenu et al. (2009) used equations for potential leaf elongation rate sensitivity to temperature (coefficients for cardinal base temperature and slope), and additionally included sensitivity to water deficit and vapor pressure deficit (VPD).

There are probably also additional environmental and nutritional factors such as solar irradiance level, N and carbohydrate supply that may also influence leaf expansion rate. We should be looking for how the individual genes code for highly specific processes (affected by enzymes, structures, etc.) and how those processes respond to those environments that lead to different emergent outcomes in different environments. If there is a G×E interaction, we believe it means we need to dig further to better understand the environmental effects relative to the action of that allele to understand why a given allele gives outcomes that differ from one environment to another.

8.6 Early Efforts to Link Genetics to Crop Models

Crop models have been used for decades to evaluate the physiological processes and genetic traits important to yield (Boote and Tollenaar 1994; Duncan et al. 1978; Elwell et al. 1987; Landivar et al. 1983a, b), but the early efforts did not have any substantial linkage to genes or genetic markers. Early crop modelers (Jones and Kiniry 1986; Wilkerson et al. 1983) used the term 'genetic coefficients' for maize and soybean models, but those coefficients were relatively few (5–15) and were generic inventions of the crop modelers. In fact, some early crop models ignored the concept of crop cultivar variation. This has changed with the advent of new genetic technologies to locate the genes, and nicely coincides with the continued development of improved crop models. As a result, a number of crop modeling groups are attempting to link to genes and markers, with the promise of optimizing genetic improvement in yield for multiple managements and environments. These dynamic process-oriented crop models are already coded for response to management and environment, and have the potential to incorporate genotypic information of genetic loci known to control multiple processes and thus predict phenotypic outcomes of growth and yield under different environments (Boote et al. 2001; White and Hoogenboom 2003; Chapman et al. 2003; Cooper et al. 2002).

The early efforts focused on linking models to genes that influenced the more predictable and easily understood crop life cycle and growth habit traits (White and Hoogenboom 1996; Hoogenboom et al. 1997, 2004; Hoogenboom and White 2003; Messina et al. 2006). These authors attempted to link alleles (presence or absence) via linear regression to the existing genotype specific parameters (GSPs) being used by the present crop modelers. One of the earliest attempts was that of White and Hoogenboom (1996) in which they linked seven genes affecting life cycle, growth habit, and seed size of common bean (*Phaseolus vulgaris* L.) to the GSPs of the GeneGro model, an early version of the dry bean model in the Decision Support System for Agrotechnology Transfer (DSSAT) software. The GeneGro model accurately predicted life cycle stages, but poorly explained yield variations across sites (Hoogenboom et al. 1997), which is not surprising given that yield is the outcome of many other genes and traits and because environmental site effects are usually more important than genetic differences (Mavromatis et al. 2002). Messina et al. (2006) followed the same approach in which they linked E-loci genes to the cultivar specific parameters of soybean, again with linear regression and alleles entered as present (1) or absent (0). The E-loci genes are basically a complex of six photoperiod-sensing genes that affect daylength-induced delay (timing) of soybean reproductive stages (Cober and Voldeng 2001). Taking this one step further, Messina et al. (2006) tested the approach against independent data on maturity date and yield measured over 5 years and 8 sites in the Illinois soybean variety trial. They used SSR (simple sequence repeats) markers to genotype cultivars in the trials to determine the E-loci and then used the gene-based coefficients to predict life cycle, growth, and yield of those cultivars. Their approach was successful, as the gene-based model, with only E loci information, was able to account for 75 % of variation for time to crop maturity and 54 % of variation in yield.

8.7 Connecting Genes in Process-Oriented Crop Models

Crop models are coded and parameterized to have responses to the environment built into the models and can be used to integrate over many processes to simulate biomass and yield accumulation over time. Chenu et al. (2009) placed the leaf expansion process sensitivities into a modified APSIM crop simulation model to evaluate integrated leaf area index (LAI), growth, and yield responses to drought environments. They connected 11 QTL markers with three 'coefficients' (temperature-sensitivity, VPD-sensitivity, and water-potential sensitivity) affecting leaf elongation rate (LER) and one coefficient affecting ASI (Anthesis-Silking-Interval, in days) of maize. The ASI effect was mimicked in the model as an effect on grain number set where tolerance (lower ASI) requires a lower threshold for assimilate supply on grain number set. They generated a hypothetical recombinant population and identified QTL combinations among 11 QTLs that were advantageous or disadvantageous to yield under multiple environments with different drought patterns. The effect of a given allele was coded similar to the way Messina et al. (2006) did it in which there is an additive effect of the allele invoked by presence (1) or absence (0). The net effect is the computation of the GSPs (in their case, a, b, and c effects on LER) as a function of the QTLs, so the crop model can proceed to simulate. The APSIM model, like other crop simulation models, generates different crop phenotypes as emergent outcomes under different management and weather conditions, because the model is coded to respond dynamically to environment and management (Hammer et al. 2010). The hourly leaf extension rate linked to the QTLs was incorporated in the APSIM model and gave rise to additional sensitivity of LAI to environmental factors (Chenu et al. 2009). Their work is a good beginning, but is only scratching the surface with a few QTLs and genes (maybe less than 5 % of the real action at the whole plant level of yield formation). In fact, an important aspect to note is that those few QTLs (genes) were placed into a model that already had many other additional 'background' genetics that were thus fixed for a medium-maturity maize hybrid. These features allowed the modelers to evaluate the individual and combined effects of those QTL markers (genes). They could just as easily put those same genes into a short-season or a long-season maize background.

Chenu et al. (2009) described pleiotropic effect on ASI-related traits, stating that the QTLs for ASI-related traits were co-located with four QTLs for leaf elongation (Vargas et al. 2006; Welcker et al. 2007). There are two possibilities: the genes may truly be 'linked' as being very close on a strand of DNA, or that a single gene results in both outcome traits somehow. A way to visualize the latter case, is that *these are the same gene or genes, but they give two 'outcomes'*, possibly because better root-water extraction, water-conservation, tolerance to high VPD (regardless of how that action may occur) will give better water-relations and that, in turn, affects leaf elongation as well as the ASI outcome. The genetic linkage possibility should be sorted out by genetic dissection of the QTL regions to attempt to separate the underlying genes for these traits. Linkage or pleiotropic effects were both positive and negative in association. As used by Chenu et al. (2009), co-localization (*potentially the same*

gene) meant that the hypothesized individual benefits could not be hypothesized, but had to be considered as linked and not separable. In our view, *these are the same genes*, and of course could not be simulated as separate effects. One of the important points, however, for crop modelers is to use caution to avoid simulating the two effects as separable, and to ensure linkage and full balance accounting of processes of water, N, and assimilate balance. In effect, leaf elongation and ASI are likely linked by water environment (plant water or turgor potential) as an influencing driver, rather than other separate drivers.

Simulated traits (greater leaf elongation or higher specific leaf area or later flowering) will increase LAI and thus act to enhance yield under well-irrigated environments, but the three traits give a similar emergent outcome (increased LAI) thus acting to reduce yield under reproductive terminal water deficit conditions (maize, Chenu et al. 2009; chickpea, Boote et al. 2013). In these three G×E cases, the common thread is the LAI effect on extent of depletion of soil water relative to use of the soil water for subsequent reproductive growth. But, the *same* alleles of the gene or genes acted to give a different outcome depending on environment! An allele may act to increase leaf area, but the integrated outcome can be positive, neutral, or negative, depending on soil water, weather, and management, which are conditions that the crop model accounts for. A gene cannot be equated directly to an emergent outcome without considering the environment. Chenu et al. (2009) reported considerable G×E interaction resulting from combinations of the QTL markers for maize in rainfed environments. These examples demonstrate the need for serious caution in use of outcome traits as the observed phenotype, and the need to focus on QTL effects on 'processes'. Parent and Tardieu (2014) compared mechanisms and simulation approaches of a number of current crop models, and concluded that processes need to be simulated at a low enough process-level whereby processes own their own temperature and environmental dependencies, such that outcomes are truly emergent properties. Nevertheless, how far should we 'drill down' to the underlying causal processes? Hammer et al. (2010) urged caution and a parsimonious approach in this regard.

The ASI effect in maize was much more important than the effect of potential leaf elongation for yield (Chenu et al. 2009). The co-location of QTLs for potential leaf elongation and ASI suggests to us that ASI per se is actually not an environmentally-stable trait, but may have other primary contributing causes. This is because a shorter ASI may be the improved water relations outcome of many other aspects such as greater water extraction, slower water use, slower leaf area expansion (to deplete less water) which in themselves may depend on more in-depth processes (as the leaf expansion already does).

The outcomes from dynamic crop simulation models can result in more complex G×E interaction effects than a purely statistical analysis of a given fixed dataset, in part because the crop models consider the biophysical limitations and feedbacks created by limiting resources such as water, light, time (temperature), etc., and because models can be used for many more multiple hypothetical or real environments. As Chenu et al. (2009) explained it, crop models provide the missing com-

ponent to allow complex G×E outcomes, and to assist in targeting genetic improvement for particular environments.

8.8 Genotype Specific Parameters in the CROPGRO Model

The CSM-CROPGRO model uses one common FORTRAN source code to simulate many different crops including legumes (Boote et al. 1998, 2002; Hoogenboom et al. 2012) and non-legumes such as tomato (Boote et al. 2012). The CSM-CROPGRO crops are simulated within the DSSAT software environment which facilitates input/output and various applications (Jones et al. 2003; Hoogenboom et al. 2012). For each crop (species), CROPGRO has a read-in species file that contains the specific parameterization, initialization, cardinal temperature relationships, etc. that cause the model to be specific for a given crop such as soybean, different from faba bean or common bean. This allows the FORTRAN code to be generic and to service many different crops. The species parameterization file is generally considered stable and fixed, and is modified only by the model developers.

In addition, the CSM-CROPGRO model has a cultivar file and an ecotype file for each crop, which allows users opportunity to mimic different cultivars. The cultivar file has 18 GSPs as listed in Table 8.1. Examples of GSPs in the model include critical short daylength, daylength-sensitivity slope, photothermal days from emergence to flowering, flowering to first pod, flowering to first seed, first seed to physiological maturity, light-saturated leaf assimilation (LFMAX), potential specific leaf area (SLAVR), flowering to end of leaf area expansion (FL-LF), seeds per pod, seed filling duration (SFDUR), pod adding duration (PODUR), seed protein concentration, and seed oil concentration. There are 16 additional GSPs in the ecotype file, but they vary infrequently and are used primarily to mimic major classes of cultivars such as determinate versus indeterminate. Useful GSPs in the ecotype file include determinacy (photothermal days from flowering to end of main stem node appearance, FL-VS), and rate of leaf appearance (TRIFOL). We anticipate the need to make additional code changes, modifying and adding cultivar, ecotype, and species coefficients as the need is demonstrated. Some species traits such as rate of root depth increase, root profile shape, N uptake per unit root, or cardinal temperatures for various processes could be moved from the species file into the ecotype or cultivar file. As evidence of cultivar variation for a trait becomes clear, the model's file of cultivar traits will grow. The challenge is to link these and future 'model' GSPs to real genes (or QTL markers). Table 8.1 illustrates large variation among dry bean, soybean, and peanut for the cultivar-specific-parameters. Dry bean, for example, has a shorter life cycle as seen in its shorter SD-PM and SFDUR parameters. Peanut, on the other hand, is insensitive to daylength, longer cycle (longer SD-PM and SFDUR), indeterminate (large FL-VS, large FL-LF, lower XFRT, and larger PODUR), and more productive (higher LFMAX), compared with bean and soybean. Cultivars within a species would tend to vary somewhat less than this example.

Table 8.1 Definitions of genotype specific parameters (GSP) for the CROPGRO model with default values for dry bean (cv. Calima), soybean (cv. MG 5 Hutcheson), and peanut (cv. Georgia Green). Two examples are given for ecotype parameters (TRIFOL, FL-VS)

GSP name	Genotype specific parameter definition	Dry bean	Soybean	Peanut
		Calima	MG 5 Hutch.	Georgia Green
CSDL	Critical short daylength below which reproductive development progresses rapidly with no daylength effect (h)	12.17	12.58	11.84
PPSEN	Slope of the relative response of development to photoperiod with time (1/h)	0.000	0.311	0.000
EM-FL	Time from emergence to first flower appearance (ptd[a])	24.8	22.0	21.2
FL-SH	Time from first flower to first pod (ptd[a])	3.0	8.0	9.2
FL-SD	Time from first flower to first seed (ptd[a])	12.0	15.5	18.8
SD-PM	Time from first seed to physiological maturity (ptd[a])	18.4	35.0	77.3
FL-VS	Time from first flower to last leaf on main axis – from ecotype file (ptd[a])	0.0	9.0	68.0
FL-LF	Time from first flower to end of leaf expansion (ptd[a])	10.0	18.0	85.0
TRIFOL	Rate of node appearance at opt. temp (node td^{-1}) – in ecotype file	0.35	0.32	0.35
LFMAX	Maximum leaf photosynthetic rate at 30 °C, 350 ppm CO_2, and high light (mg CO_2 m^2 s^{-1})	0.98	1.05	1.45
SLAVR	Specific leaf area of cultivar under standard growth conditions (cm^2 g^{-1})	305.0	400.0	270.0
SIZLF	Maximum size of full leaf (compound leaf) (cm^2)	133.0	230.0	18.0
XFRT	Maximum fraction of daily growth partitioned to seed + shell	1.00	1.00	0.95
WTPSD	Genetic potential weight per seed (g)	0.68	0.18	0.69
SFDUR	Seed filling duration for pod cohort (ptd[a])	15.0	23.0	42.0
SDPDV	Seeds per pod at standard growth conditions (# pod^{-1})	3.50	2.05	1.65
PODUR	Duration of pod addition (ptd[a])	11.0	10.0	28.0
THRSH	Threshing percentage, maximum % of seed to seed + shell	85.0	78.0	80.0
SDPRO	Potential seed protein (fraction)	0.235	0.40	0.27
SDLIP	Potential seed lipid (fraction)	0.030	0.20	0.51

[a]ptd, photothermal days account for thermal and photoperiod effects. One photothermal day occurs per calendar day, if temperature is optimum and if daylength is shorter than the critical short daylength (for short day plants)

8.9 Modeling Genotype by Environment by Management (G × E × M) Effects

Crop model simulations have shown that G × E × M interactions can be an outcome when the same genetic trait (result of genes controlling a process) gives an advantage in one environment but a disadvantage in a second environment (Boote 2011; Boote et al. 2003; Hammer et al. 2004; Hammer and Vanderlip 1989; Sinclair and Muchow 2001; Sinclair et al. 2000; Yin et al. 2000). In other words, an allele may have different effects on yield in different environments. This is strongly illustrated in Table 8.2 for simulation of genetic traits of chickpea when grown either under fully irrigated conditions or under water-limited terminal drought (Boote et al. 2013). Chickpea in India is typically sown at the end of the monsoon and depends on residual soil water on high-clay soils. Sensitivity tests of GSPs in Table 8.2 show that target environment, in this case irrigation management, was important, because the responses to given GSPs were frequently opposite and large for contrasting soil water availability, e.g., when specific leaf weight (SLW) was increased, simulated yield was about 11 % lower in irrigated but 18 % higher under water-limitation. Increased SLW had a negative effect on yield under irrigation because it reduced LAI and light interception. But the same higher SLW trait was beneficial under rainfed conditions because it reduced LAI, light interception, and transpiration (thus conserving water for subsequent grain yield later in the life cycle). Later flowering

Table 8.2 Grain yield response to variation in GSPs, simulated for 22 years for Annigeri chickpea grown under either rainfed or irrigated conditions at Patancheru, India. Sown on day 302 on a very fine montmorillonitic clay soil, starting at field capacity. Simulated with the CSM-CROPGRO chickpea model as developed by Singh and Virmani (1996) and modified by Singh et al. (2014b) (Table used with permission from Boote et al. (2013))

	Rainfed		Irrigated	
Cultivar coefficient modified	Mean yield (kg/ha)	Percent change (%)	Mean yield (kg/ha)	Percent change (%)
Standard simulation (Annigeri)	773		2614	
Rooting/SLW/LFMAX traits				
+10 %, rate of root depth progression	791	2.3	2614	0.0
+10 % leaf photosynthesis (LFMAX)	783	1.3	2961	13.3
+10 % specific leaf weight (SLW)	916	**18.5**	2328	**–10.9**
Life cycle traits				
10 % longer from emergence to anthesis	669	**–13.5**	3016	**15.4**
10 % longer seed-fill (First seed to maturity)	787	1.8	2893	10.7
Seed size/partitioning				
10 % larger potential seed size	746	**–3.5**	2709	**3.6**
10 % faster pod addition	774	0.2	2717	3.9

(which increases LAI) was very beneficial (15.4 % increase) under irrigation, but was negative (13.5 % decrease) under the terminal drought. The common factor was the amount of LAI produced and amount of soil water extracted (or left) before seed growth began.

These simulated outcomes are strongly supported by lysimeter experiments on chickpea (Anbazhagan et al. 2015) in which yield enhancement under terminal drought was related to transgenic genotypes that had earlier onset of flowering/podset and conserved water extraction during the vegetative phase, but increased use during the reproductive phase, thus leading to enhanced harvest index and yield. The transgene, rd29A: DREB1A (Dehydration Responsive Element Binding factor gene), clearly reduced water extraction rate during the vegetative phase (under adequate or deficit water), possibly via undocumented effect to reduce stomatal conductance. The authors did not report LAI, SLW, or leaf conductance, but biomass accumulation was frequently less for the transgenics, consistent with lower conductance and lower early water use. Table 8.2 shows that increasing photosynthesis (LFMAX) was unimportant (1.3 % increase) under rainfed conditions (also generally agreeing with Anbazhagan et al. 2015), but gave a large (13.3 %) increase under irrigated production. Faster root depth progression had minor impact only under rainfed conditions, resulting in 2.3 % yield increase, but the simulated soil was much deeper than the lysimeter study. Increased seed size and faster pod addition were beneficial under irrigation, but not under terminal drought. These simulated genetic traits showed differential responses depending on water environment (Singh and Virmani 1996) and as also shown in other studies (Singh et al. 2012).

Model sensitivity analysis of genetic traits has been used to design improved ideotypes of peanut for target environments in Thailand (Suriharn et al. 2011), of peanut for regions in India under climate change scenarios (Singh et al. 2012), and of chickpea for regions in Asia and Africa (Singh et al. 2014b). Singh et al. (2012) found that the benefit of varying SLW was positive or negative for peanut depending on location because rainfall and soil water were site-dependent, similar to the chickpea example above. Singh et al. (2013a) found that a 2 °C higher temperature tolerance of reproductive processes (flower fertilization, pod-addition, and partitioning) gave greater benefit to peanut yield in warm locations and under future warmer climate, but no benefit in cooler conditions. Similarly, the value of heat tolerance or drought tolerance traits (deeper rooting) in chickpea depended considerably on the temperature and rainfall/soil conditions of the targeted environments (Singh et al. 2014b). Boote et al. (2011) reported that traits of determinacy, higher SLW, and earlier pod addition were of minimal value for soybean under ambient CO_2, because those traits resulted in smaller LAI and less light interception. But those same traits had greater benefit under elevated CO_2, because CO_2 stimulation of vegetative growth enhances the LAI and light interception while allowing the other attendant benefits of those traits for higher leaf photosynthesis and longer grain-fill. Genotype-by-environment (site) interactions for peanut pod yield were simulated with the CROPGRO — Peanut model by Putto et al. (2013). Cross-over and non-cross-over G×E effects on peanut yield were found to be associated with combinations of five GSPs: seed-filling duration (SD-PM & SFDUR), determinacy of pod addition (PODUR and XFRT), and photosynthesis (LFMAX).

8.10 Using Crop Models to Evaluate Benefits of Single Genes/Traits and in Combination

Crop models have the ability to modify single processes, one trait at a time in a fixed background, to assess the effect on growth and final yield. In addition, they can simulate the degree of additivity of traits or cancelling effect of the traits (Boote et al. 2003; Hammer et al. 1996; Singh et al. 2012). The models can test the genetic traits in crops grown in different simulated environments, to determine those environments in which the traits have positive or negative effects. Traits when placed in combinations may act in an additive way, giving individual main effects, or they may potentially interact in a positive or negative manner depending on climate and management. Our first conclusion is that we find the beneficial traits tend to combine in an additive way at least for soybean, peanut, chickpea, sorghum, and maize modeling (Boote et al. 2001, 2003; Boote 2011; Singh et al. 2012). Singh et al. (2012) found that individual traits when placed into combinations of three to five traits together gave additive effects to improve peanut yield by 12–23 %, and the additivity of traits was as large or larger (15–29 %) under future climate change compared with the baseline climate. This gives some optimism that genetic improvement can contribute to adaptation to climate change. Boote et al. (2003) found that individual traits when placed in combinations for soybean were mostly additive with 15 % yield improvement easily achieved with only three example traits whether under baseline or elevated CO_2.

8.11 Phenotyping: Needed to Make the Connection to Genes

As mentioned previously, it is becoming easier, faster, and cheaper to find the genes and QTL markers as the molecular technologies advance. The larger problem from the aim of predicting performance is the need to collect phenotype data of performance observed in single and multiple environments, which can then be connected to genes. The challenge of predicting the phenotype from a given genotype is that phenotype results from the expression of the genotype as modified by the environment at any level of organization of an individual. Thus, a single genotype may have multiple phenotypes, depending on environment. Furthermore, phenotyping in a single environment is very limiting, as plant phenotype in one environment may not duplicate that in other environments. This brings up the questions of "What does phenotype mean?" and "What is a phenotype?".

8.11.1 What Is a Phenotype?

Phenotype includes characteristics of a plant that can readily be observed visually such as architectural structure, leaf number, leaf size, plant size, rooting depth, flower color, seed size, etc. But phenotype can also be extended to measurable

quantities such as plant mass, grain yield, grain protein, grain oil, leaf N concentration, etc. which may not be easily visible. Phenotype could also include measureable physiological and biochemical quantities and qualities such as photosynthesis or stomatal function or presence of a given enzyme or biochemical compound. While all these are relevant examples, one may want to distinguish between phenotypes that relate to final yield performance compared with intermediate botanical, chemical, or physiological characterization. Very often, the botanical, chemical and physiological traits are not well connected to final yield performance. We conclude that phenotyping can be highly diverse and the advice of plant breeders is needed to identify the phenotype resulting from genetics.

Because advanced genetic marker technology is rapid and allows for genotyping massive numbers of lines, there is intense interest and need for rapid and nearly automated phenotyping of large numbers of plants (Benfey and Mitchell-Olds 2008). This can be aimed at all the types of plant measurements summarized above, whether they are biochemical, enzymatic, visible, structural, or mass quantities. Measurement environments can range from laboratories to greenhouses to fields. There are a number of automated greenhouse methodologies in use by well-funded industrial and research laboratories, but our preference is for the phenotyping to be done under field environments because field plants are not well represented by measurements on individual spaced plants on automated carousels growing under limited light and artificial media in greenhouses as described by Arvidsson et al. (2011). White et al. (2012) described approaches for rapid phenotyping with various types of sensors in field environments, where sensors were transported over the field canopies by various means including tractors, cranes, drones, and satellites that have very accurate geospatial positioning. For example, plant height can be sensed directly, while LAI can be estimated by the Normalized Difference Vegetation Index (NDVI). By combining measurements of plant height, NDVI, and foliage temperature, White et al. (2012) suggested that canopy transpiration could be inferred. Data on these measurements taken over time throughout the full crop life cycle can be used in conjunction with dynamic crop models in an 'inverse modeling' approach to solve for genetic growth attributes. Measurements taken over time have the added advantage of allowing the 'inverse' characterization of abiotic and even biotic stresses on plants that occur over time in the field. We concur with Campos et al. (2004) that genetic differences in drought tolerance should only be assessed in field environments because pots, chambers, and greenhouses are unrealistic representations of field level hydrodynamics.

8.11.2 Example of Phenotyping of Dry Bean Recombinants *in a Field-Oriented Project*

In a common bean project, we phenotyped up to 200 recombinant inbred lines (RILs) of common bean for many growth and development traits that were measured under field conditions, with relatively minimal emphasis on final yield

(Clavijo-Michelangeli et al. 2013; Clavijo-Michelangeli 2014). The RI family was generated from a cross between Calima and Jamapa, cultivars belonging to Andean and Meso-American gene pools, respectively. The RILs were in the F11:14 generation cross (Vallejos et al. 2000), thus allowing us to assume that the RILs were homozygous. The RILs were grown in five environments with contrasting temperatures and photoperiod (Citra, Florida; Palmira, Colombia; Popayan, Colombia; Puerto Rico; and North Dakota).

The phenotypic traits we measured in this study were selected to represent traits typically measured in growth analyses studies as well as the type of outputs produced by mechanistic simulation models such as CSM-CROPGRO which we used in this study. The measured traits are listed in Tables 8.3 and 8.4 for non-destructive and destructive sampling, respectively. The traits consisted of phenological 'time-to' events (time to anthesis, to first pod, to first seed, etc.) along with many other traits measured over time, such as node number, leaf area, height, width, leaf size, plant mass, leaf mass, stem mass, pod mass, seed mass, seed size (Clavijo-Michelangeli 2014). The intent is to use inverse modeling to solve for GSPs from this data.

Table 8.3 Non-destructive time-to-event traits measured on bean recombinants at five sites

Name	Description	Frequency (days)
VE	Day of emergence (TD)[a]	Every 2
V1	Day of appearance of first unfolded true leaf (TD)	Every 2–3
R1	Day of first flower anthesis (PTD)[b]	Every 2–3
R3	Day of first pod (>2.0 cm) observed (PTD)	Every 2–3
LLMS	Day that last leaf expanding on main stem is observed (PTD)	Every 2–3
LLP	Day of last leaf expanding on plant (main stem or branches) (PTD)	Every 2–3
R5	Day of first fully-elongated 'rigid' pod, beginning seed (PTD)	Every 3 after R1
R7	Day when plant has at least one yellow-brown pod (R7) (PTD)	Every 2–3, after R5
LPA	Time from first pod added (R3) until last 2-cm pod is added (LPA) (PTD)	Every 2–3, after R5
R7-R8	Days from R7 to harvest maturity (R8) (PTD)	Every 2–3, after R5
PI	Plastochron index at three time points before R1 (TD)	3 times, at 5-d intervals
WTPSD	Average weight per seed (WTPSD)	At maturity
CNPYH	Canopy height	Weekly
CNPYW	Canopy width	Weekly

[a]Thermal days
[b]Photothermal days

Table 8.4 Destructive traits measured on bean recombinants at five sites

Name	Description	Frequency (calendar days)
Main stem		
LA1-5	Leaf area of successive individual leaves on nodes 1–5 of main stem	Weekly
LAMS	Leaf area of leaves on main stem (other leaves)	Weekly
DW1-5	Dry weight of first five leaves on main stem	Weekly
LWMS	Dry weight of leaves on main stem (other leaves)	Weekly
NODEM#	Number of nodes on main stem	
LGTHMS	Total length of main stem	
MSDW	Dry weight of main stem	Weekly
PDWMS	Dry weight of petioles on main stem	
PMS#	Number of pods on main stem	
Branches		
BR#	Number of branches on main stem	
LABS	Leaf area of leaves on branches	
LWBS	Dry weight of leaves on branches	
PDWBS	Dry weight of petioles on branches	
NODEB#	Node number on branches	
BLGT	Total length of branch stems	
BDW	Dry weight of branch stems	Weekly
PB#	Number of pods on branches	
Overall	**Weekly, also at final harvest**	
SHWGT	Shell dry weight	Weekly
SDWGT	Seed dry weight	Weekly
SDPDV	Observed # of seeds per pod, (10 pod sample)	Weekly, after R3 and at R8
Shelling %	Ratio of (seed weight/(seed weight + shell weight))	Weekly, after R3, and at R8
Final	**Final only**	
WTPSD	Final average weight per seed	At final harvest (R8)
SPPRO	Fraction protein in seeds (g(protein)/g(seed))	At final harvest (R8)
SDLIP	Fraction oil in seeds (g(oil)/g(seed))	At final harvest (R8)

8.11.3 Inverse Modeling *Tools for Optimizing Genetic Parameters from Phenotyping Data*

One of the goals of our common bean project is to estimate GSPs for each RIL, using the CSM-CROPGRO-Bean model in conjunction with an optimizer program. The GSPs are then to be linked to QTL markers with various genetic algorithms as described later. The parameter estimation program (Alderman 2013) is a hybrid algorithm incorporating a Gibbs-sampler (Casella and George 1992) within a version of the Metropolis-Hastings algorithm (Chib and Greenberg 1995). Essentially, the algorithm generates vectors of candidate parameter values and evaluates them

with a log-likelihood equation. With each iteration of the optimization, candidate values for parameters are generated using a normal random walk routine. With these candidate values the CROPGRO model is simulated, and then the log-likelihood for the new values of parameters are calculated. We learned by experience that a staged sequential optimization approach (Table 8.5) is much better for the optimization process than simultaneously throwing all the GSPs against all of the measured crop variables. In the staged-sequential approach, we solved appropriate sets of GSPs (parameter groups), likely to influence a set of given crop model outputs which were matched to target crop variables. The process was continued for 1000s of iterations for each parameter group, where the number of iterations depends on number of GSPs and number of target crop variables (phenotypic traits) influenced by the GSPs. One reason for the general linear flow of the staged-sequential approach in Table 8.5 is life cycle progression (where one stage must occur before we predict the next one). Another reason is that some GSPs are uniquely connected to a limited number of more stable crop target observations and those are thus solved first (seeds per pod, seed size, potential seed-to-pod ratio). Likewise, the early vegetative growth GSPs are optimized before the productivity-related GSPs such as LFMAX. There may be some iteration as well, because large change in LFMAX can cause variation in LAI which also affects canopy assimilation. Use of combinations of several GSPs against multiple targets can be used to accomplish the same goal, but require knowledge and experience in which GSPs to group against targets. We have used this approach in an automated process for each given genotype. Results of optimized GSPs obtained show promise and give good differential phenotypic outcomes of growth dynamics for different genotypes (but results are not available for this paper).

8.12 Linking Genes to GSPs in CROPGRO Legume Models

8.12.1 Example of Gene-Linked Model for CROPGRO-Soybean

The CROPGRO-Soybean model was linked to E loci and QTL markers by Messina et al. (2006), and then used to predict life cycle and grain yield in Illinois state variety trials. Phenotypic data was collected in field experiments in Florida during two seasons on near isogenic lines for genotypes that had previously been characterized for six E loci. Then an optimization approach was used with that data to solve for the GSPs of life-cycle duration for CROPGRO-Soybean. Then the obtained values of the GSPs were related, with multiple linear regression, as linear functions of the E loci, where 0 indicated recessive and 1 indicated dominance nature of a given loci. The allelic information at loci E1, E2, E3, E4, E5, and E7, respectively, were used. An additional variable NLOCI was also created to represent the number (sum) of dominant alleles at the E loci (sum ranges from 0 to 6, none dominant to all dominant), because Stewart et al. (2003) showed a linear relationship between

Table 8.5 Sequence of optimization of GSPs with the CROPGRO-bean model for genotypes of common bean (*Phaseolus vulgaris*) Jamapa×Calima recombinant inbred population grown at five sites

	Crop model		Target field measurements
Sequence	GSP name	Description	
1	PL-EM	[a]PT-time between planting and emergence	Emergence
2	EM-V1	PT-time required from emergence to first true leaf	V1
3	EM-FL	PT-time from emergence to flower appearance	R1
	PPSEN	Slope of rel. response to photoperiod, using all sites	
4	RWDTH	Canopy width (m)	Canopy height
5	RHGHT	Canopy height (m)	Canopy width
6	FL-SH	PT-time from first flower to first pod	R3
7	FL-SD	PT-time from first flower (R1) to first seed (R5)	R5 (full size pod)
8	SD-PM	PT-time from first seed (R5) to physiol. maturity (R7)	R5-R7
9	R7-R8	PT-time from physiol. (R7) to harvest maturity (R8)	R7 & 95 % mature pods
10	SLPF	Soil fertility factor, calibrated for each site *across all genotypes*	Slope of above-ground biomass at each site
11	TRIFL	Rate of appearance of leaves on main stem (MS)	Node number
12	FL-VS	PT-time from first flower to last leaf on main stem	Node #, last node on main stem
13	SDPDV	Number of seeds per pod	Seeds per pod at final harvest
14	WTPSD	Potential seed size (weight)	Average seed weight at final harvest
15	THRSH	Percentage seed in pod (%)	Shelling % at maturity
16	SLAVR	Specific leaf area ($cm^2\ g^{-1}$)	Specific leaf area over time
17	SIZLF	Leaf area of fifth fully expanded leaf (cm^2)	LAI over time, through expansion of fifth MS leaf
18[b]	LFMAX	Light-saturated leaf photosynthesis rate (mg $CO_2\ m^{-2}\ s^{-1}$)	Crop mass & pod mass over time
19	FL-LF	PT-time from first flower to end of leaf expansion on plant	Last leaf expansion on plant, or peak LAI
20	PODUR	PT-time required to reach final pod load	Time between addition of first and last pod
21[c]	FL-SD	PT-time from first flower (R1) to first seed (R5)	Onset rapid pod & seed
22[c]	SD-PM	PT-time from first seed (R5) to physiol. maturity (R7)	R7& 95 % mature pods

[a]PT represents photo-thermal time
[b]May iterate on LFMAX & SIZLF if LAI not predicted well after step 18
[c]If step 21 results in shift in life cycle (FL-SD), then SD-PM is re-solved

photoperiod sensitivity and the number of dominant alleles at E loci. One could view this almost as multiple doses of the same genetic region. Viewed in this way, the critical short daylength (CSDL, in hours) is determined by E3, E5, and the sum of dominant loci (NLOCI). The photoperiod sensitivity slope (PPSEN) is determined by E1 and NLOCI. Even though the equation for CSDL explicitly lists E3 and E5, and that for PPSEN lists E1, the presence of dominant alleles at the other loci not in the equation act via the NLOCI variable. The photothermal time from emergence to flowering (EM-FL) is determined by E1 and E3.

$$\text{CSDL} = 14.33 - 0.44\ \text{NLOCI} + 0.27\ \text{E3} - 0.48\ \text{E5} + 0.18\ \text{NLOCI E5}$$

$$\text{PPSEN} = 0.11 + 0.063\ \text{NLOCI} + 0.58\ \text{E1} - 0.13\ \text{E1 NLOCI}$$

$$\text{EMFL} = 20.77 + 2.1\ \text{E1} + 1.8\ \text{E3}$$

Other GSPs, such as photothermal time from V1 to end of juvenile (V1-JU), beginning flower to beginning seed (FL-SD), time from beginning flower to end of main stem (FL-VS), and time from beginning seed to beginning maturity (SD-PM) were similarly expressed as a function of these genes. The GSPs, parameterized in year 1 from the genes, successfully predicted flowering, onset of pod addition, last leaf formed on the main stem, and beginning maturity of lines grown in a second season. The most important test was done against independent data on phenology and yield of soybean cultivars grown in the Illinois state variety trial conducted at 8 sites over 5 years. A set of genotypes in that trial were genotyped with SSR markers to determine their respective E loci, and those E loci were used to compute the GSPs. The gene-based model, with only E loci information, was able to account for 75 % of the variation in date of maturity and 54 % of the yield variation. This indicates that life cycle is relatively easier to predict, and secondly that life cycle is an important contributor to yield potential of soybean. Mavromatis et al. (2002) also confirmed that solved GSP affecting soybean life cycle are more repeatable across environments than are GSPs affecting yield.

8.12.2 Example of Gene-Linked Model for CROPGRO-Common Bean

An example of common bean GSPs from genes defined in the CSM-GeneGro model (Hoogenboom et al. 2004) is illustrated here. Hoogenboom et al. (2004) worked with seven genes: Ppd (long daylength delay), Hr (enhances effect of Ppd, requires Ppd to be present), Fin (indeterminate), Fd (early flowering and maturity), Ssz1 (seed size), Ssz2 (seed size), and Ssz3 (seed size). The GSPs were described (similar to Messina et al. 2006) as a function of these genes.

$$\text{PPSEN} = 0.004 + 0.0154\ \text{Ppd} + 0.036\ \text{Hr} - 0.0104\ \text{Ppd}\ \text{Hr}$$

$$\text{EMFL} = 26.77 + 4.886\ \text{Fin} - 5.88\ \text{Fd}$$

$$\text{FLSH} = 4.63 + 0.972\ \text{Ssz1} - 0.98\ \text{Ssz2} - 1.8\ \text{Ssz3}$$

$$\text{FLSD} = 10.61 + 2.028\ \text{Ssz2} - 2.1\ \text{Ssz3}$$

$$\text{SDPM} = 21.027 - 0.11\ \text{Ssz1} + 4.13\ \text{Hr}$$

$$\text{FLVS} = 7.00 + 4.76\ \text{Fin} - 2.75\ \text{Ssz2} - 1.02\ \text{Fin}\ \text{Ssz2}$$

$$\text{FLLF} = 18.0 + 3.8\ \text{Fd} - 6.9\ \text{Ssz2}$$

$$\text{SLAVR} = 322 + 41\ \text{Ssz1} - 38\ \text{Ssz2} - 25.3\ \text{Ssz3}$$

$$\text{WTPSD} = 0.22 + 0.21\ \text{Ssz1} + 0.07\ \text{Ssz2}$$

$$\text{SDPDVR} = 5.14 - 0.2\ \text{Fin} - 1.9\ \text{Ssz1} + 0.24\ \text{Ssz3}$$

Other GSPs not shown were also a function of the genes. With these genes and the CROPGRO-Dry Bean model, we illustrate different modeled growth dynamics for several of the cultivars that differ in presence or absence of the dominant allele for those genes (where first-letter capitalization indicates the dominant allele). Cultivars Porrillo Sintetico (mostly Meso-American) is Ppd, hr, Fin, fd, ssz1, ssz2, and Ssz3, while cultivar Manitou is Ppd, hr, fin, Ssz1, Ssz2, Ssz3. Cultivar Bayo Madero is Ppd, Hr, Fin, Ssz1, Ssz2, Ssz3. By changing 'Fin' to 'fin', we created a Calima-like cultivar with Ppd, hr, fin, Ssz1, Ssz2, Ssz3. The emergent behavior of the genes are illustrated in Figs. 8.1, 8.2, and 8.3, and also compared to observed growth data of Porrillo Sintetico and Jatu Rong (Calima-like) cultivars measured at Palmira in 1990 by Sexton et al. (1994). Figure 8.1a illustrates how genes for time to flower as well as determinancy affect main stem node number. The Calima-like cultivar is the only one with fin, whereas the other three gene-based cultivars have dominant allele Fin, along with different photo-thermal times to flower. Figure 8.1b shows the GSP-based model simulation against data collected on Porrillo Sintetico and Jatu Rong at Palmira by Sexton et al. (1994). The simulated Calima-like gene-based cultivar closely matches the data on node number on main stem for the Jatu Rong cultivar and the simulated gene-based Porrillo Sintetico fits well with the observed data on Porrillo Sintetico. The gene-based cultivars had considerable variation in pattern of LAI (Fig. 8.2a), reaching a peak sooner for the Calima-like (fin) with early flowering (Fd), followed by the cultivars with Fin but relatively mid-cycle flowering, and later peak with highest LAI for gene-based Porrillo Sintetico. Comparison with real data for Jatu Rong and Porrillo Sintetico at Palmira (Fig. 8.2b), illustrates close predictions of LAI patterns for the gene-based Calima-like and Porrillo Sintetico. Lastly, there are different total seed growth patterns for the four gene-based cultivars (Fig. 8.3a), with earliest onset and lowest yield associated with the Calima-like

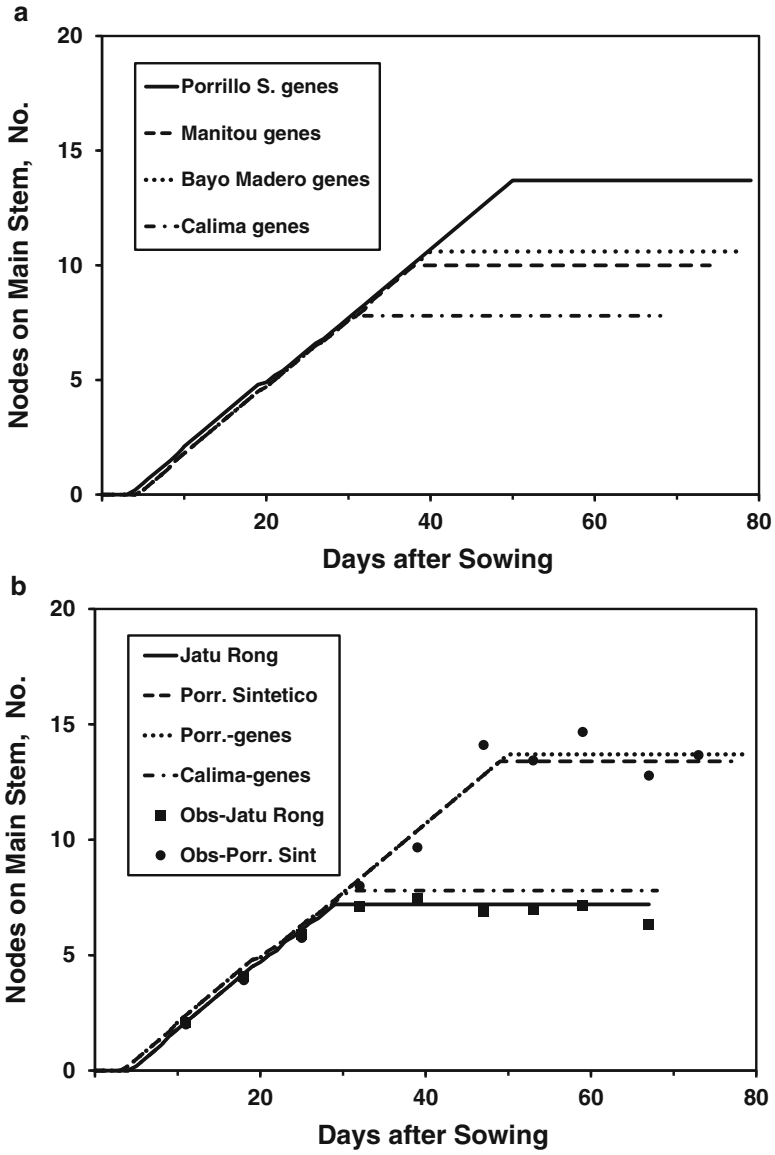


Fig. 8.1 (**a**) Node number on main stem as affected by genes in four gene-based dry bean cultivars at Palmira, (**b**) Node number on main stem as affected by genes in two gene-based dry bean cultivars compared to data for Jatu Rong (Calima-like) and Porrillo Sintetico at Palmira

cultivar, followed by later onset and higher seed yield for the other gene-based cultivars. Comparison with observed data on seed mass over time for Jatu Rong and Porrillo Sintetico at Palmira (Fig. 8.3b) shows good and comparable predictions of seed growth patterns for the gene-based Calima-like and Porrillo-Sintetico cultivars. The point of these comparisons is to illustrate the resulting widely varying

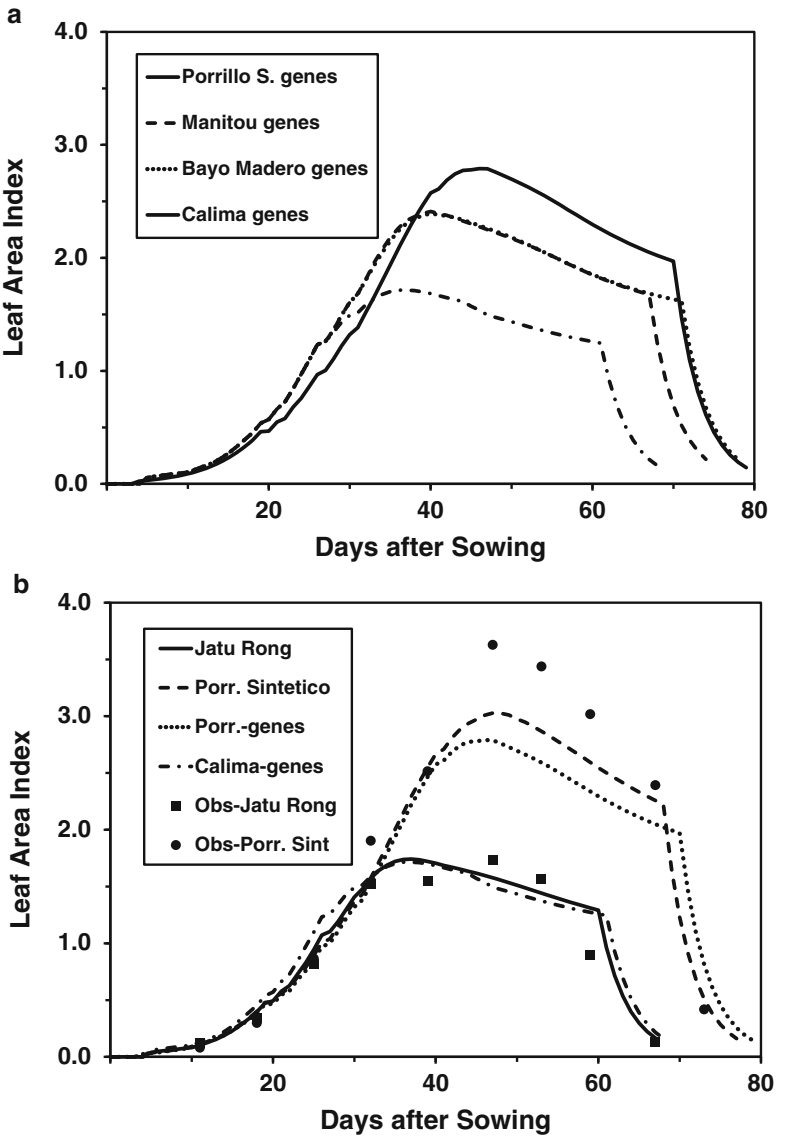


Fig. 8.2 (**a**) Leaf area index as affected by genes in four gene-based dry bean cultivars at Palmira, (**b**) Leaf area index as affected by genes in two gene-based dry bean cultivars compared to data for Jatu Rong (Calima-like) and Porrillo Sintetico at Palmira

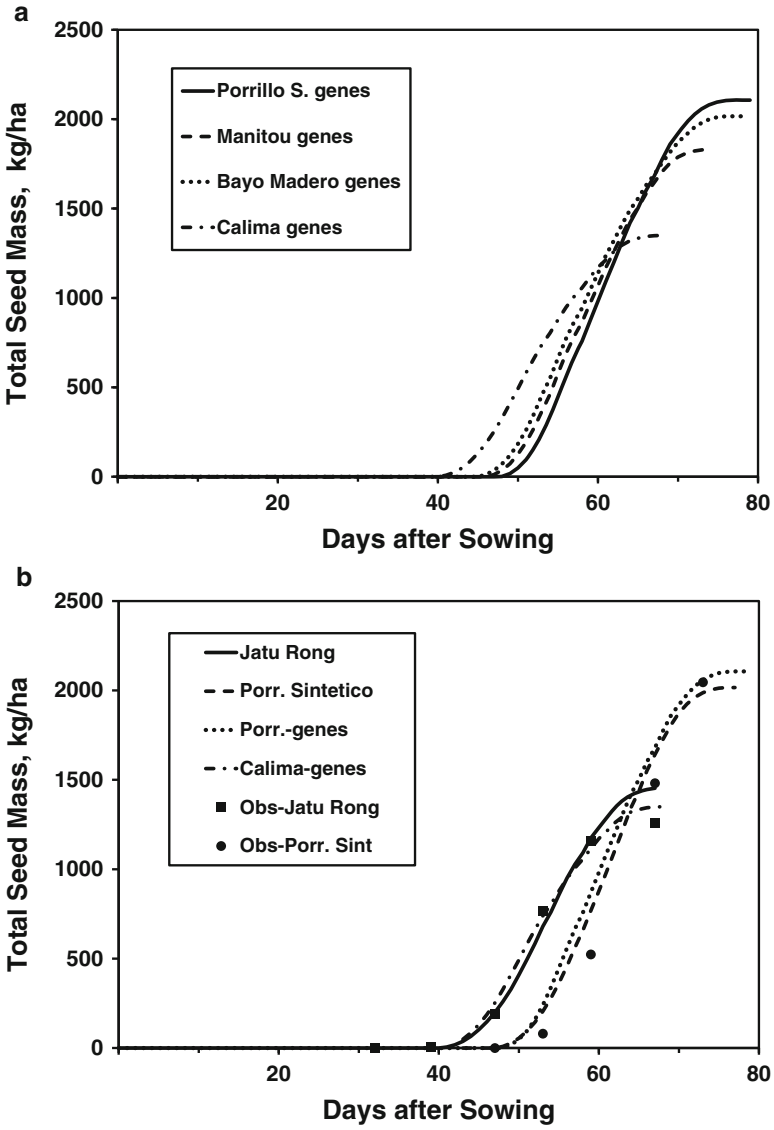


Fig. 8.3 (**a**) Seed mass over time as affected by genes in four gene-based dry bean cultivars at Palmira, (**b**) Seed growth pattern as affected by genes in two gene-based dry bean cultivars compared to data for Jatu Rong (Calima-like) and Porrillo Sintetico at Palmira

growth patterns attributed to genes that control life cycle progression and other aspects of dry bean growth.

8.13 Conclusions

There appears to be considerable potential for linking genetics to mechanistic process-oriented crop models, with the goal of predicting field phenotypic performance as a function of genes. However, there remain a number of issues to fully take advantage of the considerable potential that this approach provides. Present crop models will need to be re-designed such that the processes in the models can be better linked to genetics. Current GSPs need to be re-considered and additional ones added. The proposed linkage of model GSPs to genes will work via the approaches used by several modelers. Present statistical analyses of genes and G×E interactions are good tools to discover important genes and the G×E interactions, but once highlighted, we propose a follow-up fruitful linkage of those genes to GSPs in crop models to more effectively predict phenotypes in multiple new environments, taking advantage of the built-in environmental sensitivities of the crop models. We hypothesize that many cases of genotype by environment interaction are often the case of one gene (or the same complex of genes), that gives a performance advantage in one environment but a negative response in another environment. The proposed approach of linking genes to crop models should lead to improved characterization of variations among cultivars in response to both environment and to genetic control of those responses to environment.

Our group at the University of Florida recently conducted a workshop in 2015 bringing together plant breeders and crop modelers to discuss QTL mapping, phenotyping, optimization of GSPs from collected phenotype data, followed by linking of GSPs in crop models to the genes (www.conference.ifas.ufl.edu/MergingCMG). The goal was cross training and interdisciplinary learning, including exercises with real data, because we feel the disciplines have diverged so much that geneticists are not well connected with the field phenotyping, and crop modelers are not connected with the geneticists.

Acknowledgements This work was supported in part by an award from the Plant Genome Research Program of the National Science Foundation (DBI/PGRP-0923975), and by the University of Florida Agricultural Experiment Station.

References

Alderman PD (2013) Improving soil fertility management in northern Ghana: an integrated modeling approach. Ph.D. dissertation, University of Florida

Anbazhagan K, Bhatnagar-Mathur P, Sharma KK, Baddam R, Kavi-Kishor PB, Vadex V (2015) Changes in timing of water uptake and phenology favours yield gain in terminal water stressed chickpea AtDREB1A transgenics. Funct Plant Biol 42:84–94

Arvidsson S, Pérez-Rodríguez P, Mueller-Roeber B (2011) A growth phenotyping pipeline for Arabidopsis thaliana integrating image analysis and rosette area modeling for robust quantification of genotype effects. New Phytol 191:895–907. doi.org/10.1111/j.1469-8137.2011.03756.x

Benfey PN, Mitchell-Olds T (2008) From genotype to phenotype: systems biology meets natural variation. Science 25:495–497

Boote KJ (2011) Improving soybean cultivars for adaptation to climate change and climate variability. In: Yadav SS, Redden RJ, Hatfield JL, Lotze-Campen H, Hall AE (eds) Crop adaptation to climate change. Wiley-Blackwell, West Sussex, pp 370–395

Boote KJ, Tollenaar T (1994) Modeling genetic yield potential. In: Boote KJ, Bennett JM, Sinclair TR, Paulsen GM (eds) Physiology and determination of crop yield. ASA/CSSA/SSSA, Madison, pp 533–565

Boote KJ, Jones JW, Hoogenboom G, Pickering NB (1998) The CROPGRO model for grain legumes. In: Tsuji GY, Hoogenboom G, Thornton PK (eds) Understanding options for agricultural production. Kluwer, Dordrecht, pp 99–128

Boote KJ, Kropff MJ, Bindraban PS (2001) Physiology and modelling of traits in crop plants: implications for genetic improvement. Agric Sys 70:395–420

Boote KJ, Mínguez MI, Sau F (2002) Adapting the CROPGRO legume model to simulate growth of faba bean. Agron J 94:743–756

Boote KJ, Jones JW, Batchelor WD, Nafziger ED, Myers O (2003) Genetic coefficients in the CROPGRO-soybean model: links to field performance and genomics. Agron J 95:32–51

Boote KJ, Rybak MR, Scholberg JMS, Jones JW (2012) Improving the CROPGRO-tomato model for predicting growth and yield response to temperature. HortSci 47:1038–1049

Boote KJ, Jones JW, White JW, Asseng S, Lizaso JI (2013) Putting mechanisms into crop production models. Plant Cell Environ 36:1658–1672

Campos H, Cooper M, Habben JE, Edmeades GO, Schussler JR (2004) Improving drought tolerance in maize: a view from industry. Field Crop Res 90:19–34

Casella G, George EI (1992) Explaining the Gibbs sampler. Am Stat 46:167–174

Chapman SC, Cooper M, Podlich D, Hammer GL (2003) Evaluating plant breeding strategies by simulating gene action and dryland environment effects. Agron J 95:99–113

Chenu K, Chapman SC, Tardieu F, McLean G, Welcker C, Hammer GL (2009) Simulating the yield impacts of organ-level quantitative trait loci associated with drought response in maize: a "gene-to-phenotype" modeling approach. Genetics 183:1507–1523

Chib S, Greenberg E (1995) Understanding the metropolis-hastings algorithm. Am Stat 49:327–335

Clavijo-Michelangeli JA (2014) Modeling reproductive growth and development in the common bean (*Phaseolus vulgaris* L.). Ph.D. dissertation, University of Florida

Clavijo-Michelangeli JA, Bhakta M, Gezan SA, Boote KJ, Vallejos CE (2013) From flower to seed: identifying phenological markers and reliable growth functions to model reproductive development in the common bean (*Phaseolus vulgaris* L.). Plant Cell Environ 36:2046–2058

Cober ER, Voldeng HD (2001) A new soybean maturity and photoperiod-sensitivity locus linked to E1 and T. Crop Sci 41:698–701

Cooper M, Podlich DW, Micallef KP, Smith OS, Jensen NM, Kruger NL (2002) Complexity, quantitative traits and plant breeding: a role for simulation modelling in the genetic improvement of crops. In: Kang MS (ed) Quantitative genetics, genomics, and plant breeding. CABI, Wallingford, pp 143–166

Duncan WG, McCloud DE, McGraw RL, Boote KJ (1978) Physiological aspects of peanut yield improvement. Crop Sci 18:1015–1020

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchel SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6(5):e19379

Elwell DL, Curry RB, Keener ME (1987) Determination of potential yield-limiting factors of soybeans using SOYMODE/OARDC. Agric Syst 24:221–242
Furbank RT, Tester M (2011) Phenomics – technologies to relieve the phenotypic bottleneck. Trends Plant Sci 16:635–644
Hammer GL, Vanderlip RL (1989) Genotype-by-environment interaction in grain-sorghum. 3. Modeling the impact in field environments. Crop Sci 29:385–391
Hammer GL, Butler DG, Muchow RC, Meinke H (1996) Integrating physiological understanding and plant breeding via crop modelling and optimization. In: Plant adaptation and crop improvement. CAB International, Wallingford, pp 419–441
Hammer GL, Sinclair TR, Chapman SC, van Oosterom E (2004) On systems thinking, systems biology, and the in silico plant. Plant Physiol 134:909–911
Hammer GL, van Oosterom E, McLean G, Chapman SC, Broad I, Harland P, Muchow RC (2010) Adapting APSIM to model the physiology and genetics of complex adaptive traits in field crops. J Exp Bot 61:2185–2202
Hoogenboom G, White JW (2003) Improving physiological assumptions of simulation models by using gene-based approaches. Agron J 95:82–89
Hoogenboom G, White JW, Acosta Gallegos J, Gaudiel RG, Myers JR, Silbernagel MJ (1997) Evaluation of a crop simulation model that incorporates gene action. Agron J 89:613–620
Hoogenboom G, White JW, Messina CD (2004) From genome to crop: integration through simulation modeling. Field Crop Res 90:145–163
Hoogenboom G, Jones JW, Wilkens PW, Porter CH, Boote KJ, Hunt LA, Singh U, Lizaso JI, White JW, Uryasev O, Royce FS, Ogoshi R, Gijsman AJ, Tsuji GY, Koo J (2012) Decision support system for agrotechnology transfer (DSSAT) version 4.5 [CD-ROM]. University of Hawaii, Honolulu
Johannsen W (1911) The genotype conception of heredity. Am Nat 45:129–159
Jones CA, Kiniry JR (1986) CERES–Maize: a simulation model of maize growth and development. Texas A&M University Press, College Station
Jones JW, Keating BA, Porter CH (2001) Approaches to modular model development. Agric Syst 70:421–443
Jones JW, Hoogenboom G, Porter CH, Boote KJ, Batchelor WD, Hunt LA, Wilkens PW, Singh U, Gijsman AJ, Ritchie JT (2003) The DSSAT cropping system model. Eur J Agron 18:235–265
Kao C-H, Zeng Z-B, Teasdale RD (1999) Multiple interval mapping for quantitative trait loci. Genetics 152:1203–1216
Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199
Landivar JA, Baker DN, Jenkins JN (1983a) Application of GOSSYM to genetic feasibility studies. I. Analyses of fruit abscission and yield in okra-leaf cottons. Crop Sci 23:497–504
Landivar JA, Baker DN, Jenkins JN (1983b) Application of GOSSYM to genetic feasibility studies. II. Analyses of increasing photosynthesis, specific leaf weight and longevity of leaves in cotton. Crop Sci 23:504–510
Lester DR, Ross JJ, Davies PJ, Reid JB (1997) Mendel's stem length gene (*Le*) encodes a gibberellin 3 beta-hydroxylase. Plant Cell 9:1435–1443
Li Y, Wu RL (2010) Functional mapping of growth and development. Biol Rev 85:207–216
Ma CX, Casella G, Wu RL (2002) Functional mapping of quantitative trait loci underlying the character process: a theoretical framework. Genetics 161:1751–1762
Malosetti M, Ribaut J-M, van Eeuwijk FA (2013) The statistical analysis of multi-environment data: modeling genotype-by environmental interaction and it genetic basis. Front Physiol 4:44
Martin DN, Proebsting WM, Hedden P (1997) Mendel's dwarfing gene: cDNAs from the Le alleles and function of the expressed proteins. Proc Natl Acad Sci U S A 94:8907–8911
Mavromatis T, Boote KJ, Jones JW, Wilkerson GG, Hoogenboom G (2002) Repeatability of model genetic coefficients derived from soybean performance trials across different states. Crop Sci 42:76–89
Mendel G (1866) Experiments in plant hybridization. Translated by The Royal Horticultural Society of London in 1938. Harvard University Press, Cambridge, MA

Messina CD, Jones JW, Boote KJ, Vallejos CE (2006) A gene-based model to simulate soybean development and yield responses to environment. Crop Sci 46:456–466
Parent B, Tardieu F (2014) Can current crop models be used in the phenotyping era for predicting the genetic variability of yield of plants subjected to drought or high temperature. J Exp Bot 65:6179–6189. doi:10.1093/jxb/eru223
Putto C, Pathanothai A, Jogloy S, Boote KJ, Hoogenboom G (2013) Determination of plant traits that affect genotype × location (G × L) interaction in peanut using the CSM-CROPGRO-peanut model. Int J Plant Prod 7(3):537–568
Reymond M, Muller B, Tardieu F (2004) Dealing with the genotype x environment interaction via a modeling approach: a comparison of QTLs of maize leaf length or width with QTLs of model parameters. J Exp Bot 55:2461–2472
Sax K (1923) The association of size differences with seed-coat pattern and pigmentation in Phaseolus vulgaris. Genetics 8:552–560
Schlötterer C (2004) The evolution of molecular markers – just a matter of fashion? Nat Rev Genet 5:63–69
Sexton PJ, White JW, Boote KJ (1994) Yield-determining processes in relation to cultivar seed size of common bean. Crop Sci 34:84–91
Sinclair TR, Muchow RC (2001) System analysis of plant traits to increase grain yield on limited water supplies. Agron J 93:263–270
Sinclair TR, Pinter PJ, Kimball BA, Adamsen FJ, LaMorte RL, Wall GW, Hunsaker DJ, Adam N, Brooks TJ, Garcia RL, Thompson T, Leavitt S, Matthias A (2000) Leaf nitrogen concentration of wheat subjected to elevated CO2 and either water or N deficits. Agr Ecosyst Environ 79:53–60
Singh P, Virmani SM (1996) Modeling growth and yield of chickpea (*Cicer arietinum* L). Field Crop Res 46:41–59
Singh P, Boote KJ, Kumar U, Srinivas K, Nigam SN, Jones JW (2012) Evaluation of genetic traits for improving productivity and adaptation of groundnut to climate change in India. J Agron Crop Sci 198:399–413
Singh P, Nedumaran S, Ntare BR, Boote KJ, Singh NP, Srinivas K, Bantilan MCS (2014a) Potential benefits of drought and heat tolerance in groundnut for adaptation to climate change in India and West Africa. Mitig Adapt Strateg Glob Chang 19:509–529. doi:10.1007/s1027-012-9446-7
Singh P, Nedumaran S, Boote KJ, Gaur PM, Srinivas K, Bantilan MCS (2014b) Climate change impacts and potential benefits of drought and heat tolerance in chickpea in South Asia and East Africa. Eur J Agron 52:123–137
Stewart DW, Cober ER, Bernard RL (2003) Modeling genetic effects on the photothermal response of soybean phenological development. Agron J 95:65–70
Suriharn B, Patanothai A, Boote KJ, Hoogenboom G (2011) Designing a peanut ideotype for a target environment using the CSM-CROPGRO-peanut model. Crop Sci 51:1887–1902
Vallejos CE, Malandro JJ, Sheehy K, Zimmermann MJ (2000) Detection and cloning of expressed sequences linked to a target gene. Theor Appl Genet 101:1109–1113
Vargas M, Van Eeuwijk FA, Crossa J, JRibaut JM (2006) Mapping QTLs and QTL x environment interaction for CIMMYT maize drought stress program using factorial regression and partial least squares methods. Theor Appl Genet 112:1009–1023
Wang Z, Pang X, Wu W, Wang J, Wang Z, Wu R (2013) Modeling phenotypic plasticity in growth trajectories: a statistical framework. Evolution 68:81–91
Welcker C, Boussuge B, Bencivenni C, Ribaut JM, Tardieu F (2007) Are source and sink strengths genetically linked in maize plants subjected to water deficit? A QTL study of the responses of leaf growth and of anthesis–silking interval to water deficit. J Exp Bot 58:339–349
White JW, Hoogenboom G (1996) Simulating effects of genes for physiological traits in a process-oriented crop model. Agron J 88:416–422
White JW, Hoogenboom G (2003) Gene-based approaches to crop simulation: past experiences and future opportunities. Agron J 95:52–64
White JW, Andrade-Sanchez P, Gore MA, Bronson KF, Coffelt TA, Conley MM, Feldmann KA, French AN, Heun JT, Hunsaker DJ, Jenks MA, Kimball BA, Roth RL, Strand RJ, Thorp KR,

Wall GW, Wang G (2012) Field-based phenomics for plant genetics research. Field Crop Res 133:101–112

Wilczek AM, Roe JL, Knapp MC, Cooper MD, Lopez-Gallego C, Martin LJ, Muir CD, Sim S, Walker A, Anderson J, Egan JF, 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

Wilkerson GG, Jones JW, Boote KJ, Ingram KT, Mishoe JW (1983) Modeling soybean growth for crop management. Trans ASAE 26:63–73

Wu RL, Lin M (2006) Functional mapping—how to map and study the genetic architecture of dynamic complex traits. Nat Rev Genet 7:229–237

Xu X, Zeng L, Tao Y, Vuong T, Wan J, Boerma R, Noe J, Li Z, Finnerty S, Pathan SM, Shannon JG, Nguyen HT (2013) Pinpointing genes underlying the quantitative trait loci for root-knot nematode resistance in palaeopolyploid soybean by whole genome resequencing. Proc Natl Acad Sci U S A 110:13469–13474

Yin X, Struik PC (2010) Modelling the crop: from system dynamics to systems biology. J Exp Bot 61:2171–2183

Yin X, Kropff MJ, Goudriaan J, Stam P (2000) A model analysis of yield differences among recombinant inbred lines in barley. Agron J 92:114–120

Zeng Z-B (1993) Theoretical basis for separation of multiple linked gene effects in mapping of quantitative trait loci. Proc Natl Acad Sci U S A 90:10972–10976

Zeng Z-B (1994) Precision mapping of quantitative trait loci. Genetics 136:1457–1468

Zhao W, Zhu J, Gallo-Meagher M, Wu RL (2004) A unified statistical model for functional mapping of genotype×environment interactions for ontogenetic development. Genetics 168:1751–1762

Chapter 9
Modelling QTL-Trait-Crop Relationships: Past Experiences and Future Prospects

Xinyou Yin, Paul C. Struik, Junfei Gu, and Huaqi Wang

Abstract Ecophysiological crop models have long been used to understand crop responses to environmental factors and to crop management practices, by integrating quantitative functional relationships for various physiological processes. In view of the potential added value of robust crop modelling to classical quantitative genetics, model-input parameters are increasingly considered to represent 'genetic coefficients', which are environment-independent and amenable to selection. Likewise, modern molecular genetics can enhance applications of ecophysiological modelling in breeding design by elucidating the genetic basis of model-input parameters. A number of case studies, in which the effects of quantitative trait loci (QTL) have been incorporated into existing ecophysiological models to replace model-input parameters, have shown promise of using these QTL-based models in analysing genotype-phenotype relationships of more complex crop traits. In this chapter, we will review recent research achievements and express our opinions on perspectives for QTL-based modelling of genotype-by-environment interactions and even epistasis on complex traits at crop level.

9.1 Introduction

A major challenge in field- and greenhouse-crop production today is breeding for genotypes and realizing their potential in given (often stressful) environments to produce sufficient high-quality products while maintaining the sustainability of production systems and resource use. This goal can be achieved via creating

X. Yin (✉) • P.C. Struik
Department of Plant Sciences, Centre for Crop Systems Analysis, Wageningen University, P.O. Box 430, 6700 AK Wageningen, The Netherlands
e-mail: Xinyou.Yin@wur.nl

J. Gu
Key Laboratory of Crop Genetics and Physiology of Jiangsu Province, Yangzhou University, Yangzhou, Jiangsu, China

H. Wang
Plant Breeding and Genetics, China Agricultural University, 100193 Beijing, China

X. Yin, P.C. Struik (eds.), *Crop Systems Biology*,
DOI 10.1007/978-3-319-20562-5_9

phenotypes of complex traits at the level of the crop – the community of mutually interacting plants, usually of the same species. A thorough insight into gene-trait-crop relationships is therefore crucial. Currently, there is an increasing recognition amongst geneticists and breeders (e.g., Tuberosa and Salvi 2006; Dwivedi et al. 2007; Langridge and Fleury 2011; Messina et al. 2011) and physiologists (e.g., Chenu et al. 2009; Zhu et al. 2011) of immediate need for physiological and computational tools to assist breeders in more effectively analysing, interpreting, translating, and integrating the outputs from high-throughput genomics research, and to help resolving genotype-by-environment interactions (G×E) efficiently and selecting the best technology interventions and associated breeding systems for their target traits and target environments.

Actually, decades ago, process-based physiological models of crop growth have already been suggested to be useful tools in supporting breeding (e.g., Loomis et al. 1979; Spitters and Schapendonk 1990). These models quantify causality between relevant physiological processes and responses of these processes to environmental variables, and, therefore, allow predictions of crop yields not restricted to the environments in which the model parameters have been derived. Crop models require environmental inputs (i.e., weather variables and management options) and physiological inputs. The latter inputs are used as model parameters for characterizing genotypic differences. These genotype-specific parameters, acronymised as GSP by Boote et al. (see Chap. 8 of this book), are also referred to as '*genetic coefficients*' (White and Hoogenboom 1996; Mavromatis et al. 2001) or 'model-input traits' (Yin et al. 2000a), implying that model-input parameters might be (at least partly) under genetic control. As model parameters can reflect certain genetic characteristics, crop modelling has long been considered a useful computational tool to assist breeding (Loomis et al. 1979; Spitters and Schapendonk 1990; Boote et al. 2001). Shorter et al. (1991) already proposed collaborative efforts between breeders, physiologists and modellers, using models as a framework to integrate physiology with breeding.

Given the common experience that crop models based on physiologically sound mechanisms can quantify and integrate responses of crop yield to both genetic and environmental factors, crop physiologists, breeders and modellers have explored the potential of using crop models in various aspects of breeding. These activities include: (1) identifying main yield-determining traits, both under poor and conducive environments for crop growth (Semenov and Halford 2009; Yin et al. 2000b; Heuvelink et al. 2007), (2) defining optimum selection environments in order to maximize selection progress (Aggarwal et al. 1997), (3) optimizing single trait values (Boote and Tollenaar 1994; Setter et al. 1995; Yin et al. 1997), (4) designing ideotypes in which trade-offs between conflicting crop traits are properly evaluated (Spitters and Schapendonk 1990; Penning de Vries 1991; Dingkuhn et al. 1993; Kropff et al. 1995; Haverkort and Kooman 1997), and (5) assisting multi-location testing (Dua et al. 1990) and explaining G×E (Mavromatis et al. 2001; van Eeuwijk et al. 2005; Bertin et al. 2010). Some of these explorative activities were summarised by Boote et al. (2001).

However, crop physiology has not contributed much to breeding (Jackson et al. 1996). All the above-mentioned studies, based on model simulations, are to give suggestions that breeders may use. Stam (1998) and Koornneef and Stam (2001), from a geneticist's and breeder's point of view, expressed their concerns about this model-based approach that ignores the inheritance of the model-input traits. For example, for designing ideotypes by modelling, it is assumed, either tacitly or explicitly, that these traits can be combined at will in a single genotype. Such an assumption ignores the possible existence of constraints, feedback mechanisms and correlations among the traits. Constraints might be imposed simply by the fact that little genetic variation exists in the genetic material available for selection. Thus, models may not identify those traits for which gain via breeding may be easiest (Jackson et al. 1996). Correlations between the traits, due either to a tight linkage between genes or to a single gene that affects multiple traits (pleiotropy), may seriously hamper the realization of an ideotype (for example, an early-maturing potato cultivar with high resistance against late blight; Visker et al. 2003; Struik 2010). Knowledge of the genetic basis of phenotypic variation, even described in terms of model-input traits, is crucial for a successful breeding programme (Stam 1998). Therefore, understanding the inheritance of the model parameters within the framework is required (Stam 1998). To assist the development of efficient breeding strategies, crop modelling requires quantitative understanding of the inheritance of the model-input parameters.

Largely to that end, there have been a growing number of studies that combine crop modelling with modern genetic approaches. In this chapter, we review recent research experiences on elucidating the QTL-trait-crop relationships by integrating crop systems modelling and genetic QTL mapping. Other roles of crop modelling in genetics and breeding will also be explored. Future prospects in this research line are discussed in the context of assisting crop improvement programmes.

9.2 Complementarity of Crop Modelling and Genetic Mapping

In genetics, complex crop traits can be unravelled into the effects of individual QTL – quantitative trait loci (Paterson et al. 1988), commonly using the materials of a segregating population derived from a bi-parental cross. A common result of QTL analysis of complex crop traits is that QTL expression is usually conditional on the environment and this greatly impedes the application of QTL-mapping information for manipulating complex traits (Stratton 1998).

Crop models can potentially be of help in this respect to better address genotype-phenotype relationships, provided that model-input parameters can be easily measured (Yin et al. 2004) and vary little with environmental conditions (Reymond et al. 2003; Tardieu 2003). Model-input parameters (or 'genetic coefficients') reflect effects of genetic origin in the way that one set of parameters represents one

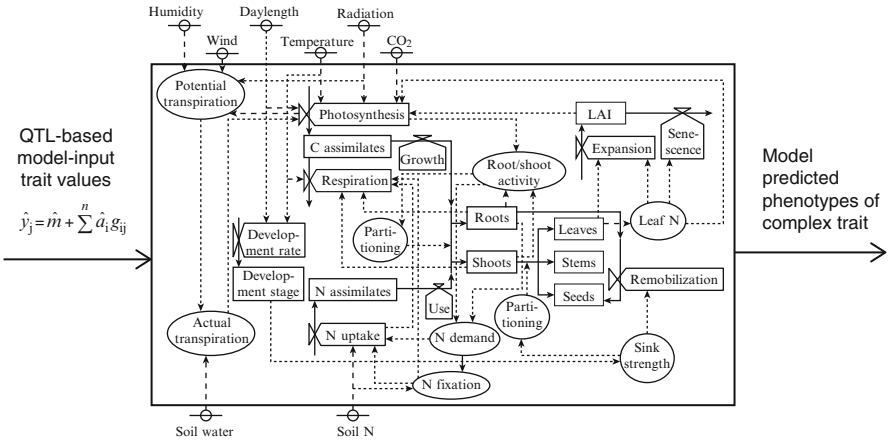


Fig. 9.1 Illustration of QTL(quantitative trait loci)-based crop modelling to predict complex phenotypes. Values of model-input traits were estimated from the effects of their identified QTL to replace the original values of the traits. QTL-based trait values $\hat{y}_j$ were calculated using a simplest genetic model, in which $\hat{m}$ is the estimated intercept, g_{ij} is the genetic (additive effect) predictor of the *i*-th QTL genotype for the *j*-th individual of the mapping population, $\hat{a}_i$ is the estimated additive effect of the *i*-th QTL on the trait ($i = 1, 2, \ldots, n$). This figure predominantly illustrates the structure of a typical crop model using the Forrester's symbols. In such a crop growth model, climatic factors (e.g., radiation, temperature, …), soil conditions, and crop parameters are inputs, and model simulation gives output of complex traits such as crop yield

genotype (Tardieu 2003). Hence, the models manifest that the crop phenotype is achieved through nonlinear interactive and ontogenetic responses of component processes to multiple environmental factors. Such an approach has added value to classical genetics, since geneticists often ignore or overlook competition, density, nutrient supply, morphology, physiology and plasticity, lumping such matters vaguely under the 'G×E' term or introducing simple response functions in their statistical models (e.g., van Eeuwijk et al. 2005).

To enhance applications of ecophysiological modelling in genetic analysis and breeding, understanding the genetic basis of model-input parameters is essential. Yin et al. (1999a, b) first showed that the QTL approach can be applied to model-input parameters to elucidate their inheritance. Such attempts have resulted in an integrated approach of so-called 'QTL-based ecophysiological modelling' (Fig. 9.1), which links crop modelling with genetics, focusing on the G×E problem and genotype-phenotype relationships. The QTL-based models can be used to predict performance of any genotype in any environment.

This approach of QTL-based modelling was first illustrated to predict a very complex trait – the grain yield of barley (*Hordeum vulgare*) by Yin et al. (1999a, b, 2000a). The same approach for QTL-based modelling analyses was applied to crop traits such as leaf elongation rate in maize (Reymond et al. 2003) and flowering time in barley (Yin et al. 2005), rice (Nakagawa et al. 2005) and *Brassica oleracea* (Uptmoor et al. 2008, 2012), fruit quality in peach (Quilot et al. 2004, 2005) and

tomato (Bertin et al. 2010; Prudent et al. 2011), maize kernel number (Amelong et al. 2015), and Arabidopsis stomatal conductance (Reuning et al. 2015), all relatively simple traits with well-defined influences of some dominant environmental factors (such as vapour pressure deficit, soil moisture content, temperature and photoperiod). In the domain of morphological traits, the phenotypic effects of QTL for culm length, grain number, and grain size have been simulated using morphologically explicit models in barley (Buck-Sorlin 2002) and in rice (Xu et al. 2011; see Chap. 2 of this book by Xu and Buck-Sorlin). These studies on relatively simple developmental, morphology-related or growth traits demonstrate that the approach can unravel G×E, and highlight the potential to analyse more complex traits manifested through season-long growth dynamics (see Gu et al. 2014b).

In short, genetic mapping dissects a quantitative trait into various genetic factors – QTL (Paterson et al. 1988), but it can only predict the trait phenotype in independent new environmental conditions to a limited extent (Stratton 1998). Ecophysiological modelling can reveal how G×E comes about (Tardieu 2003), but it does not consider the genetic basis of model parameters that describe genotypic differences. Combining ecophysiological modelling and genetic mapping can dissect complex traits into component traits, integrate effects of QTL of the component traits over time and space at the whole-crop level, and predict yield performance of various genetic make-ups under different environmental conditions.

9.3 Roles of Modelling in Assisting Genetic Analysis and Breeding

There is *in silico* evidence that this combined modelling and genetic approach can facilitate translating the QTL mapping into more efficient marker-assisted breeding strategies (Hammer et al. 2006). To this end, more accurate crop models would facilitate the improvement of efficiencies of combined model- and marker-assisted breeding. In this section we summarise, in abroad sense, the applications of modelling in support of genetic analysis and breeding programme.

9.3.1 Models Can Support Phenotyping for the QTL Mapping

A pre-requisite of the proper use of phenotypic data for quantitative genetic analysis is that the phenotypic data of the different genotypes should be collected under the same environmental conditions and at the same plant developmental stage. On the other hand, quantitative genetic analysis requires screening of a large population to realize the required genetic resolution based on high power of the analyses. Complicated statistical analyses and experimental designs were often used to remove environmental errors, for example, caused by heterogeneity in the

experimental field. But for highly sensitive traits (such as photosynthesis), microclimate fluctuations could also obscure the genetic effects existing in the population. Ecophysiological models based on solid physiological knowledge could be useful tools to standardize the measurements (Gu et al. 2012a). Using model-based standardization, several QTL related to photosynthesis were found under fluctuating field conditions, and were confirmed in independent greenhouse environments. Ecophysiological models can thus play a role in improving the quality of data on traits that are sensitive to environmental changes.

Another example was reported by Yin et al. (1999a), who mapped specific leaf area (SLA) in a barley recombinant inbred lines population. After adjusting SLA values measured at the same chronological time to values at the same physiological age, the effect on SLA from the *denso* gene was no longer significant. The effect of the *denso* gene detected at the same chronological time was therefore the consequence of its direct effect on flowering time. An ecophysiological model can thus indeed assist QTL analysis by removing either environmental noise or indirect effects from other traits.

Breeders often have a crude method of phenotyping. Modelling can help to upgrade their phenotyping activities. Khan (2012) used several expressions to describe phasic development curves of canopy cover dynamics in potato. Not only the overall area under the curve but also individual model parameters were found to vary among individuals of a mapping population, and the parameters most related to the area under the curve were identified, providing the trait components selectable for improving canopy light interception and biomass yield.

9.3.2 Models Can Dissect Complex Traits into Physiological Components

Physiological modelling can dissect complex traits (e.g., photosynthesis or yield) into physiological component traits. Gu et al. (2012b) used a photosynthesis model to dissect photosynthesis into: (1) stomatal conductance g_s, (2) mesophyll conductance g_m, and (3) electron transport capacity J_{max} and Rubisco carboxylation capacity V_{cmax}. Using the crop growth model GECROS, yield was connected to, and dissected into seven physiological input parameters (Gu et al. 2014b). By dissecting complex traits into physiologically meaningful component traits, it is possible to assess genetic variation for each component trait and evaluate its relative importance by sensitive analyses or regression analyses. For example, genetic variation in light-saturated photosynthesis and transpiration efficiency was found to be mainly associated with variation in g_s and g_m (Gu et al. 2012b). The physiological input trait 'total crop nitrogen uptake at maturity' was found to have the most significant effect on yield (Gu et al. 2014b). Similarly, Prudent et al. (2011) combining an

ecophysiological modelling and QTL analysis, identified key elementary processes and genetic factors underlying tomato fruit sugar concentration. All these results show that the physiological model could be helpful to decide on priority targets for breeding, although possible impact remains to be validated through actual breeding and field testing.

9.3.3 Models Can Integrate and Project Single Organ Level Genetic Variation to Crop Level

Modelling not only can dissect complex traits into physiological relevant components, but can also integrate effects of QTL of the component traits over time and space, and predict complex traits at the whole-crop level of various genetic make-ups under different environmental conditions (Yin and Struik 2010). This could be useful to evaluate the effect of changes in a single trait or single trait-related QTL on a crop, while keeping other traits constant to avoid the confounding effects from other physiological processes, which is not plausible in a 'real' experiment. For example, as stated earlier, improving photosynthesis is generally thought crucial for improving plant production, but often no correlation or even negative correlations between photosynthesis and plant production were observed (Evans and Dunstone 1970; Teng et al. 2004; Zhao et al. 2008; Jahn et al. 2011; Gu et al. 2014b). The reason for this discrepancy could be that plants differed genetically in many respects other than photosynthesis. Hence, Gu et al. (2014a) used the crop model GECROS, and found that the natural genetic variation in leaf photosynthesis within our experimental mapping population would result in equivalent differences in production when scaled up to crop level. The ability of integration and upscaling can also help evaluate impacts of QTL for a specific organ-level trait at crop level in a different environment. Using the CROPGRO-soybean model, Messina et al. (2006) estimated the effects of QTL markers from a set of near-isogenic lines and satisfactorily predicted the variation of yield across five years and eight sites among an independent set of soybean cultivars. Chenu et al. (2009), using the crop model APSIM-Maize, evaluated a QTL accelerating leaf elongation on maize yield. This QTL could cause a yield increase in an environment with water deficit before flowering, but reduced yield under terminal drought stress. This information could be used in breeding for specific environments or for facing the challenges caused by climate change. Most importantly, the feature of integration could allow for designing ideotypes of various genetic make-ups underlying physiological processes. Based on the genetic variation and resulting QTL for each physiological component in photosynthesis, it was shown that the ideotype for leaf-level photosynthesis and transpiration efficiency (TE) could potentially be improved by 17.0 % and 25.1 %, respectively (Gu et al. 2012b).

9.3.4 *Ecophysiological Model May Help to Resolve Genetic Complexities*

A simple genetic model can be assumed for QTL analysis of the component traits, but more sophisticated genetic control (G×E, and epistasis, i.e., the interaction between genes) on the complex trait per se can be manifested when QTL-based parameter values are fed-back to the ecophysiological model. As discussed earlier, use of ecophysiological models to predict and interpret G×E has been widely recognised and exemplified (Reymond et al. 2003; Yin et al. 2005; Messina et al. 2006; see Chaps. 3, 4, 5, 6, 7, and 8 of this book). The use of the models to interpret epistasis is less recognised. Epistasis is often found for phenotypes that are achieved through interactive and interrelated metabolic and ontogenetic pathways (Lee 1995). It might be reduced or even disappear if input traits of a model that accounts for interrelations among relevant processes are subjected to analysis. Such possibility agrees with the awareness of geneticists that epistasis can often be removed by a physiologically based scaling of trait values (Kearsey and Pooni 1996). For example, crop yield is analysed in agronomy as the product of several yield component traits; independent QTL on various yield components must exhibit an epistatic effect on yield (Yin et al. 2002). Chapman et al. (2003) used the crop model APSIM to generate a state space of genotype performance based on 15 genes controlling 4 traits and then search this space for selection. They showed complex epistatic and G×E effects were generated for yield even though gene action at the trait level had been defined as simple additive effects. Similarly, White and Hoogenboom (1996), Messina et al. (2006), and White et al. (2008) used simple linear additive models to regress model-input parameters against several known gene loci across cultivars or genotypes, implicitly modelling the epistatic effects of these genes on the aggregated traits such as yield or days to flowering. It should be acknowledged that use of crop models to resolve epistasis in real experimental populations may be a more difficult task than to resolve G×E, and for the required accuracy crop models should evolve into crop systems biology models (see Chap. 1 of this book by Baldazzi et al.).

9.3.5 *QTL-Based Modelling Can Quantify Constraints in Breeding*

Model simulation could inspire breeders. However, Stam (1998) and Koornneef and Stam (2001), from a geneticist's perspective, expressed their concerns that the ignorance of the inheritance of the model-input parameters is a major constraint for breeders to adopt the results of model-based approaches. Often in ideotype design by modelling, modellers implicitly assumed that plant traits can be combined at will

into a single genotype. As stated earlier, such an unrealistic practice ignores the possible existence of constraints, feedback mechanisms and correlations among traits. By integrating crop modelling with genetics – QTL-based modelling, it is possible to evaluate constraints in breeding either due to limited genetic variation or to correlations. Gu et al. (2012b) showed trade-offs between improving photosynthesis and TE either due to tight linkage or to pleiotropic effects of QTL related to g_m and g_s. If the linkage between g_m and g_s, or co-location of QTL of g_m and g_s could be broken, the virtual ideotype could have both improved photosynthesis and TE. The quantitative importance of breaking this linkage could be used together with insights of geneticists about chances of success in guiding decisions in breeding programmes, thus strengthening the scientific basis for designing breeding activities.

9.3.6 QTL-Based Modelling Can Assist Marker-Assisted Selection

Marker-assisted selection (MAS), combined with conventional breeding approaches, has been used to effectively integrate major genes or QTL with large effect into widely grown varieties (Jena and Mackill 2008). The use of cost-effective DNA markers and a MAS strategy will provide opportunities for breeders to develop high-yielding, stress-tolerant, and better quality rice cultivars. For example, pyramiding different resistance genes using MAS provided opportunities to breeders to develop broad-spectrum resistance against diseases and insects (Huang et al. 1997). An example of the latter approach is the insertion of cassettes of up to four resistance genes from wild potato species into existing cultivars using cisgenesis techniques to make these existing cultivars resistant to late blight (Haverkort et al. 2009). By stacking several resistance genes, the resistance cannot be broken easily by the causal agent *Phytophthora infestans*, certainly not when this approach is combined with a well-designed resistance management strategy (Haverkort et al. 2009).

Gu et al. (2014b) also showed that the existing GECROS model can be a useful tool to enhance marker-assisted breeding through a model-based ideotype design. Using the principles for QTL-based modelling as defined earlier (Yin et al. 2000a, 2004, 2005), marker-based crop modelling was performed to rank the markers identified for various yield-determining physiological traits that are input parameters of GECROS (Table 9.1). It was found that the relative importance of markers differed markedly between well-watered and drought-stressed environments (the correlation coefficient in the rank between the two environments was 0.09; $P > 0.10$). Such an analysis detected markers that breeders can prioritize in their MAS programmes for specific environments. Gu et al. (2014b) showed that compared with identification of markers through multiple regression for yield per se, the model-based approach identified additional QTL and could be complementary to the analysis of yield per se.

Table 9.1 Percentage of the phenotypic variation in yield among rice introgression lines (ILs) (R^2) accounted for by different sets of simulations using the marker-based version of the GECROS model, when markers were fixed one at a time to calculate different sets of marker-based parameter inputs for GECROS. Marker positions are denoted as 'Chr_cM', that is, Chromosome_centiMorgan, as identified during QTL-analysis (Based on Gu et al. (2014b))

Fixed marker		Well-watered		Drought-stressed	
Chr_cM	Name	R^2 (%)	Rank	R^2 (%)	Rank
1_9.5	RM8068	51.6	11	42.6	13
1_25.4	RM8145	53.9	18	41.4	10
1_98.1	RM306	51.6	11	39.5	6
1_124.8	RM1152	50.9	9	44.6	18
2_92.5	RM475[a]	46.2	3	37.7	5
2_110.9	RM1367	51.7	14	45.5	19
2_139.3	RM8030[a]	34.2	1	40.9	9
3_79.1	RM251	47.9	5	46.2	20
3_108.4	RM338[b]	52.6	17	29.8	1
4_25.5	RM518	59.4	20	44.1	17
4_123.8	RM2799	51.8	16	40.2	7
5_20.6	RM7302[b]	51.6	11	33.2	2
7_43.5	RM432	50.7	8	36.9	4
7_47	RM11	51.7	14	43.8	16
7_81.05	RM3753	49.4	7	41.9	11
8_83.7	RM284[a]	45.7	2	42.6	13
9_0.8	RM5799	48.3	6	42.6	13
9_64.4	RM410[a, b]	47.3	4	35.9	3
10_87.1	RM294A	51.1	10	40.5	8
12_61.6	RM1261	53.9	19	42.1	12
Baseline simulation		51.6		42.6	

The baseline simulation gives the R^2 values for the simulation, in which no marker was fixed, i.e., IL-specific allelic values (−1 or 1) were used for all markers in calculating marker-based inputs; for other sets of simulations, markers were fixed one at a time, in which all ILs were assumed to carry an identical allele (i.e., 0) at the locus of the considered marker in calculating marker-based inputs

[a]These markers were also identified for yield per se under well-watered conditions

[b]These markers were also identified for yield per se under drought-stressed conditions

9.4 Past Experiences in Integrating Ecophysiological Modelling and Genetic Mapping

The main purpose of practising QTL-based ecophysiological modelling on the basis of using a mapping population is to predict genotypic impact on phenotypes, in contrast to conventional crop modelling which usually aims to predict the impact of environmental variables on crop productivity. The following contains summaries of current experiences.

9.4.1 *Models Generally Perform Better in Simulating Phenotypic Differences Caused by Environmental Variation than by Genotypic Differences*

This may not be surprising given that when individuals of mapping population are phenotyped in multiple environments, it is common that variance due to environmental differences is much more significant than the variance due to genotypic differences (Yin et al. 2000a, b; Gu et al. 2012b; Khan et al. 2014). Crop models generally perform well in assessing the impact of significant environmental variation due to changes in climatic variables and nutrient availabilities, as most existing crop models were built to predict the impact of environmental variables on crop productivity. It is a challenge to predict the impact of a subtle change in traits among relatively similar lines within a breeding population. Therefore, model's suitability in analysing genotype-to-phenotype relationships in an experimental population needs critical examinations (Parent and Tardieu 2014). The following two aspects deserve particular attentions. First, better modelling of the final spikelet or seed number of cereals under stress conditions is needed as stress sensitivity of this sink-size trait often shows larger genetic variability than that of the source-activity traits. Second, the genetic difference in response to soil environments is currently subject to huge uncertainty, due partly to the lack of sufficient site-specific information about the soil and partly to uncertainties in modelling root growth and soil processes (Gu et al. 2014b; Khan et al. 2014). There is an obvious need for robust algorithms for rooting density and depth for resource capture from the soil and their genotypic variabilities.

9.4.2 *Some Model-Input Parameters Do Not Contribute to the Model in Explaining Differences among the Genotypes*

The importance of model parameters in contributing to explaining yield differences among individuals of mapping population can be evaluated by fixing them once at a time at their average value (Yin et al. 2000b; Khan et al. 2014). It is expected that the model explained percentage of yield differences will drop if the parameter fixed is important for yield determination. Counter-intuitively, fixing some parameters, which seem to be physiologically important, even increased the explained percentage of phenotypic variation. Identified examples for such parameters are: specific leaf area (SLA), leaf nitrogen content, post-flowering duration for barley (Yin et al. 2000b), vegetative growth period and maximum plant height for potato (Khan et al. 2014). Similarly, when introducing genetic variation of individual biochemical parameters of leaf-photosynthesis into the photosynthesis sub-model of GECROS, the variation of yield accounted for by GECROS decreased significantly for both well-watered and drought-stressed conditions (Gu et al. 2014b). Such model based sensitivity analysis suggests whether or not the model has incorporated right

parameters in explaining yield differences among genotypes in a population. The reasons for the unimportance of those seemingly important parameters in terms of yield physiology in explaining genetic differences among genotypes remain to be elucidated.

9.4.3 Some Model-Input Parameters Are Hard to Measure for the Whole Population, and Heritability Estimates (h^2) of Model-Input Parameters Are Generally Lower than Those of Classical Plant Traits

Some model-input parameters are used in crop models in a tabular form, e.g., coefficients for assimilate partitioning among growing organs in Wageningen SUCROS-family models. Determining values of these coefficients require frequent destructive samplings during growing season, which can be implemented in classical agronomic experiments but are practically infeasible for individual lines of a mapping population. These types of parameters certainly do not allow high throughput measurements, and many of them need many steps to measure. Measurement noise accumulates over steps; some parameters require curve-fitting method to estimate, which again involves some fitting uncertainty/noise (also see Chap. 5 by Luquet et al.). So, the h^2 for measured phenotypic data of these parameters is often lower than for traits relating to classical agronomic, plant size, and architecture traits. This is in analogy to the result of Jahn et al. (2011) that physiological traits such as stomatal conductance (which involve various steps of measurements and calculations) had a lower h^2 than the classical agronomic and morphological traits. As a consequence, the percentage of phenotypic variation explained (r^2) by QTL identified for model parameters is often lower than the r^2 of QTL for classical plant traits if measured in the same experiments (Yin et al. 1999b; Gu et al. 2014b).

9.4.4 The Percentage of Phenotypic Variation of a Complex Trait Accounted for by the QTL-Based Model Is Comparable with, or Slightly Lower Than, That Obtained from the Original Parameter Values

QTL-based model parameters values derived from QTL-statistics can partly remove the noise of phenotypic values of model-input parameters. On the other hand, QTL identified by mapping analysis account for only part of genetic variance of parameters. So, a common feature of these studies is that predictability of QTL-based models is nearly comparable with that of the model using original parameter values (see Fig. 9.2 for an example), as the gain from removing random noise in original

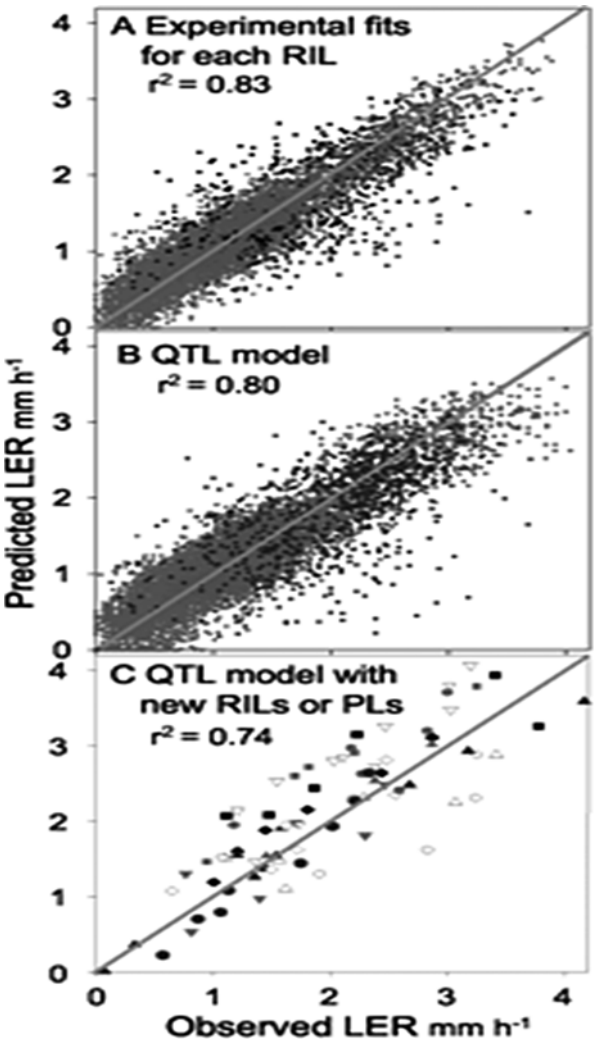


Fig. 9.2 Comparison between observed maize leaf elongation rates LER and those fitted by a simple ecophysiological model for LER (*A*), simulated by the model using QTL-based model input parameters (*B*), and QTL-based model predicted LER for those recombinant inbred lines RILs that were not included for QTL analysis (C) (Redrawn from Reymond et al. 2003)

parameters by QTL-statistics is roughly cancelled out by the loss due to the fact that the identified QTL cannot explain 100 % of the genetic variance of the parameter values (e.g., Yin et al. 2000a, 2005; Reymond et al. 2003; Gu et al. 2014b). However, QTL-based models can predict the performance of genotypes that were not phenotyped for model-input parameters (Fig. 9.2c) as long as marker data at or near QTL are available for these genotypes.

9.4.5 *Despite the Outnumbering of QTL for Model-Input Parameters Relative to Those for the Complex Trait per Se, the Percentage of Phenotypic Variation of a Complex Trait Accounted for by Its QTL of Model-Input Parameters Is Lower than That Obtained from Complex-Trait QTL*

Current QTL statistics can hardly find more than eight QTL for a quantitative trait to avoid false positives in QTL analysis (Kearsey and Farquhar 1998), although this may greatly depend on the population size. Therefore, rather than looking for QTL for a complex trait itself, determining QTL for underlying component traits might give more information. Indeed, using crop models to dissect a complex trait into its individual components will help to identify more QTL than analysing the complex trait per se (Gu et al. 2014b; Amelong et al. 2015). However, the percentage of phenotypic variation of a complex trait accounted for by its QTL of model-input parameters is lower than that obtained from complex-trait QTL (Prudent et al. 2011; Yin et al. 2000a; Gu et al. 2014b). Low predictability of the models could be the reason for that. However, even when yield is dissected into yield components using a simple arithmetic formula (typically: yield is equal to the product of yield component traits) and the formula perfectly predicts yield variation, the percentage of phenotypic variation of yield accounted for by QTL of its component traits is lower than that obtained from complex-trait QTL (Yin et al. 2002; Fig. 9.3). This suggests that phenotyping of model-input parameters and yield component traits may involve more random noise.

9.4.6 *Number of QTL Identified for Model-Input Parameters Based on a Bi-parental Population Is Limited; Most Model-Input Parameters Are Often Affected by the Pleiotropic Effect of 1–2 Major QTL*

Dominance of a major QTL is a common phenomenon, presumably due to the contrast between the parents intentionally chosen in making the bi-parental mapping population. Typically, one parent represents a modern cultivar that is currently widely cultivated whereas the other is an old traditional genotype that was probably cultivated before the Green Revolution. This means that one or two major genes are segregating in the population. Very often major genes not only affect major morphological characteristics and yield level, but also have pleiotropic effects on multiple phenological and physiological traits including model-input parameters. This has

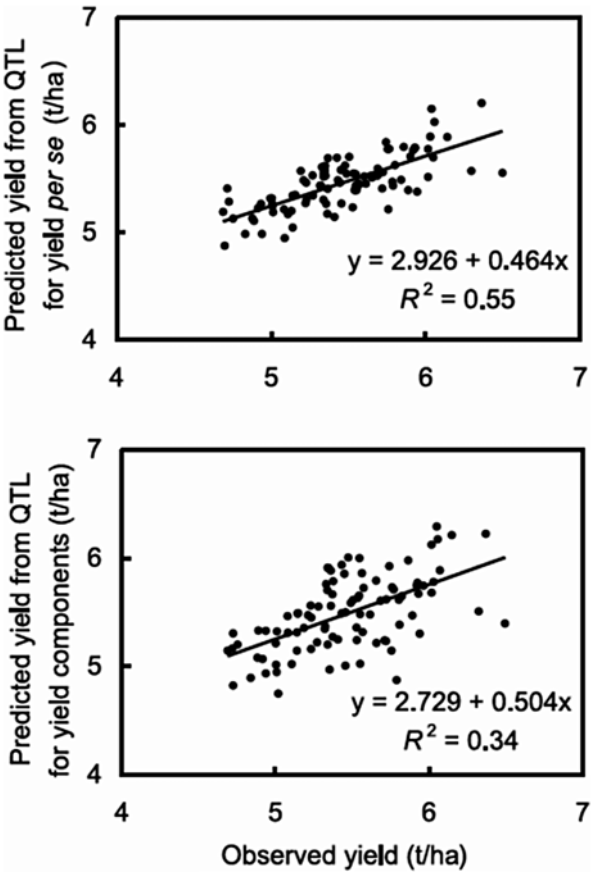


Fig. 9.3 Comparison between observed values of grain yield and those predicted from quantitative trait loci (QTL) identified for yield itself, and between observed values of grain yield and those predicted from QTL identified for its three component traits: spikes per m^2, number of kernels per spike and 1000-kernel weight (Redrawn from Yin et al. 2002)

been shown by, for example, the *denso* gene in barley (Yin et al. 1999a, b), *rht* genes in wheat (Baenziger et al. 2004; Laperche et al. 2006; White 2006), the maturity-class gene on chromosome V in potato (Khan 2012), and the RM410 locus on chromosome 9 in rice (Gu et al. 2012a). Of these, the *denso* gene in barley is particularly pleiotropic and its dominant effect is ubiquitous, not only on plant height, yield and yield components (Yin et al. 1999b), but also for flowering parameters (Yin et al. 2005) and traits like SLA (Yin et al. 1999a) and nitrogen use efficiency (Kindu et al. 2014).

9.4.7 *Medium-Size Population Is the Best Option That Combines Feasibility and Robustness in Integrated Ecophysiological Modelling and Genetic Mapping Studies*

From a statistical point of view, the larger is the population, the more robust is QTL mapping (Vales et al. 2005). Going for a large population size is generally not feasible because most individual input parameters in existing crop models do not allow high-throughput phenotyping. Because of the cost and/or time needed, researchers often went for selective phenotyping, and some were even pushed to phenotype only 46 individuals as a subset of a population to identify QTL for model-input parameters (Uptmoor et al. 2012), thereby, greatly sacrificing the statistical power of QTL detection. In that sense, crop models should be improved in a way that most parameters would be measurable in phenotyping facilities (Yin et al. 2004; Parent and Tardieu 2014). Before such a model becomes available, a medium-size population consisting of ca 100 individuals as a comprise of phenotyping feasibility and QTL-detection robustness, may be the best option, if model input-parameters can be measured with the currently available methods. If model input-parameters are hard to measure, it is better to use an introgression line (IL) population other than populations like recombinant inbred lines (RILs), doubled-haploid lines (DHs), as ILs differ in a lower number of loci. For example, leaf photosynthesis is commonly measured by gas exchange that does not allow high-throughput phenotyping. In a study where the entire light- and CO_2-response curves needed to be phenotyped via gas exchange, Gu et al. (2012b) chose 13 ILs (including parents) and did succeed to localize the genomic regions for seven parameters of a biochemical photosynthesis model.

9.5 Future Prospects

9.5.1 *Understanding Physiological Basis of QTL and Genetic Variation*

From a physiologists' point of view, a logic step following the mapping of genetic basis (i.e., QTL analysis) of a physiological trait is to elucidate the deeper-level physiological basis of the detected QTL underlying its genetic variation. Few studies have investigated the physiological basis of QTL underlying genetic variation of quantitative traits. The physiological basis of QTL may best be elucidated with physiological models that dissect complex traits into individual component traits. This was recently reported for leaf photosynthesis. QTL for light saturated leaf photosynthesis (A_{max}) and other related traits were first identified using an introgression line population (Gu et al. 2012a). To elucidate the physiological basis of these

QTL, combined gas exchange and chlorophyll fluorescence data were collected for entire CO_2 and light response curves of leaf photosynthesis (A), with which biochemical and physiological parameters of a combined conductance-biochemical photosynthesis model were estimated. Because measuring entire response curves is time consuming and does not allow high throughput, 13 lines (including the two parents) were carefully selected as representatives of the population, based on the QTL for leaf photosynthesis reported by Gu et al. (2012a). The curves were assessed at two stages (flowering and grain filling) for plants grown under moderate drought and well-watered conditions (Gu et al. 2012b). Using these curves, photosynthesis was then quantitatively dissected into three different physiologically relevant component traits: (1) stomatal conductance (g_s), (2) mesophyll conductance (g_m), and (3) biochemical efficiency and capacity. Although the effects of development stage and water supply on photosynthesis were predominant, significant genetic variation in the three mentioned component traits was found. Genomic regions of the variation of these biochemical parameters of photosynthesis were localised (Fig. 9.4). Genetic variation in A_{max} and TE (transpiration efficiency) was mainly caused by variation in g_s and g_m, which suggests more efforts should be focused on g_s and g_m in breeding programmes for improving photosynthesis and TE. Gu et al. (2012b) showed that relationships between these photosynthetic parameters and leaf nitrogen or dry matter per unit area, which were previously found across environmental treatments, were also valid for variation across genotypes. Therefore, they speculated that variation in photosynthesis due to environmental conditions and the variation in photosynthesis due to genetic variation within the same environment may share common physiological mechanisms.

Gu et al. (2012b) next used the model to evaluate the potential of utilizing the genetic variation in these components for improving photosynthesis (A) and transpiration efficiency (TE). Based on the genetic variation of physiological components underlying A and TE, ideotypes were designed by combining alleles positively influencing different components of photosynthesis. Model calculations showed that these ideotypes can potentially improve photosynthesis and TE significantly, compared with the best genotype of the 13 lines investigated. It was shown that if the tight link between g_m and g_s could be broken, both photosynthesis and TE could be improved simultaneously, despite the common negative correlation between A and TE (e.g., Condon et al. 2004). This result would be especially interesting for breeding for semi-arid environments.

The importance of mesophyll conductance in improving leaf photosynthesis has also been identified for materials of other genetic backgrounds in rice (Adachi et al. 2013). Adachi et al. (2014) further indicated that high leaf nitrogen content and high hydraulic conductivity are two additional physiological mechanisms contributing to high leaf photosynthesis of their near-isogenic lines (NILs), which differ from the recipient parent in only one or two introgression regions of previously mapped QTL and therefore best suit for elucidation of physiological basis for individual QTL. Similar results have been found for the genetic variation in leaf photosynthesis across cultivars in rice (Taylaran et al. 2011; Lauteri et al. 2014) and wheat (Jahan et al. 2014).

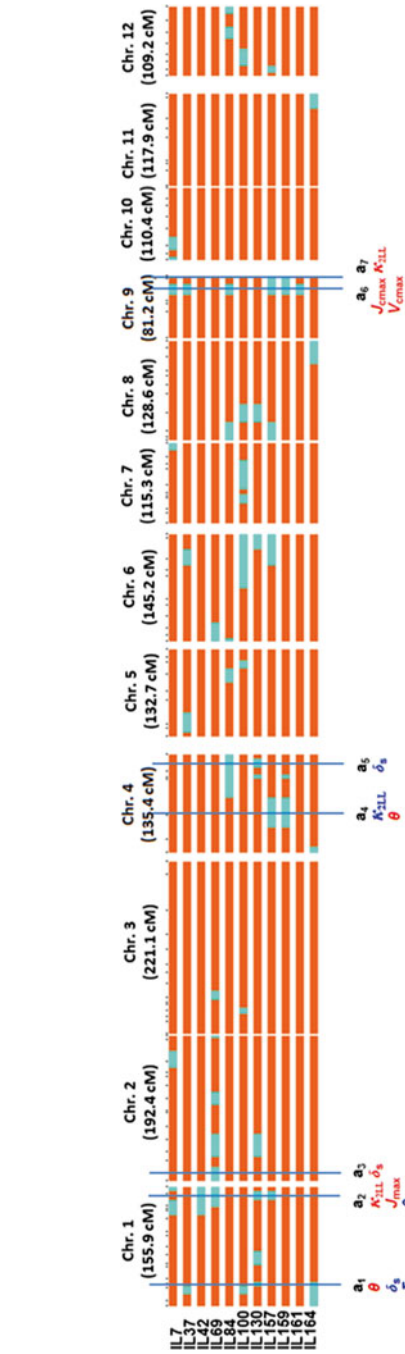


Fig. 9.4 Graphical genotypes for the 11 introgression lines which were used in the previous analysis of Gu et al. (2012b) for assigning the genome regions (QTL) to six photosynthesis parameters in a biochemical model: κ_{2LL} (efficiency of converting incident light into linear electron transport under limiting light), J_{max} (maximum electron transport rate under saturating light), θ (convexity factor of light response curve of electron transport), δ_m (parameter associated with mesophyll conductance), δ_s (parameter associated with stomatal conductance), and V_{cmax} (maximum capacity of Rubisco carboxylation). The length of each linkage group is shown in centiMorgan (cM). The *light-blue* regions indicate the introgression regions from the donor parent 'Haogelao'; the *ochre* backgrounds indicate the homozygous regions from the recurrent parent 'Shennong265'. Parameters, on which genome alleles from 'Haogelao' have positive effects and negative effects, are shown in *red* and *blue colours*, respectively (Redrawn from Gu et al. (2012b))

For more complex traits rather than leaf photosynthesis, Gu et al. (2014b) used the crop model GECROS to dissect yield into seven phenological, morphological and physiological parameters. It was found that nitrogen uptake, grain nitrogen concentration and pre- and post-flowering durations are important, whereas leaf photosynthesis was surprisingly not important, in explaining yield differences among the individual lines within a genetic population. Using rice NILs that harbour one or two spikelet-number QTL, Ohsumi et al. (2011) found that the NILs having increased spikelet number per panicle did not greatly increase grain yield because of compensation between different yield components. They also showed that the slight yield advantage of the NILs harbouring double QTL relative to other genotypes was associated with higher translocation of carbohydrates from reserves to panicle. These indicate a pleiotropic effect of the spikelet-number QTL on other physiological traits.

We call for more studies on elucidating physiological basis of QTL and pleiotropic effect of the QTL on other physiological traits/processes. Such information will facilitate to improve existing crop models that better capture physiological processes and parameters related to genetic variation of crop yield.

9.5.2 Broadening Genetic Background of the Mapping Population

While the proposed QTL-based modelling approach could potentially deal with G×E, it cannot solve all limiting factors, especially not the non-transferability of information obtained from one cross to another. The non-transferability can be largely due to the possibility that a QTL detected in one cross does simply not segregate in a second cross because the parents of the second cross carry identical alleles at that QTL – the lack of allelic diversity within a mapping population. A gene 'important' for physiologists or modellers might be useless for geneticists or breeders because if the gene is physiologically crucial, its variation will have been strongly reduced over generations of breeding (Prioul et al. 1997); so QTL will hardly be detected at such a gene locus. In this context, the approach as practised for a bi-parental cross should be extended in future although such extended studies with a broader genetic background can be most feasibly applied to simple traits that can be scored by high-throughput phenotyping.

9.5.2.1 Use of Multiple Mapping Population

With the crop model GECROS, Gu et al. (2014b) used the marker-based parameter values derived from a population of 94 introgression lines to simulate yield variation among 251 recombinant inbred lines of the same parents in rice. More directly working with multiple populations, Welcker et al. (2011) have compared the genetic

architectures of the sensitivities of maize (*Zea mays*) leaf elongation rate with evaporative demand and soil water deficit as quantified in a simple ecophysiological model. The former was measured via the response to leaf-to-air vapour pressure deficit in well-watered plants, the latter via the response to soil water potential in the absence of evaporative demand. Genetic analyses of each sensitivity were performed over 21 independent experiments with (1) three mapping populations, with temperate or tropical materials, (2) one population resulting from the introgression of a tropical drought-tolerant line in a temperate line, and (3) two introgression libraries genetically independent from mapping populations. A very large genetic variability was observed for both sensitivities. Some lines maintained leaf elongation at very high evaporative demand or water deficit, while others stopped elongation in mild conditions. A complex architecture arose from analyses of mapping populations, with 19 major meta-QTL involving strong effects and/or more than one mapping population. A total of 68 % of those QTL affected sensitivities to both evaporative demand and soil water deficit. In introgressed lines, 73 % of the tested genomic regions affected both sensitivities. They demonstrated that hydraulic processes, which drive the response to evaporative demand, also have a large contribution to the genetic variability of plant growth under water deficit in a wide range of genetic material comprising of multiple populations.

On the genetic side, geneticists are trying to improve QTL mapping resolution with several generations of intercrossing when establishing the RIL population, e.g. advanced intercross RILs. Meanwhile allelic diversity within a mapping population can be increased by intercrossing multiple genetically diverse genotypes before establishing the RILs, e.g., MAGIC – the Multi-parent Advanced Generation InterCross (Huang et al. 2011).

9.5.2.2 Genome-Wide Association Study

Virk et al. (1996) empirically showed that quantitative variation of many agronomic traits in the rice germplasm is associated with allelic variation of DNA markers, indicating that marker-trait associations not only may be present in segregating populations, but can also be manifest across a germplasm collection of a crop species. Later studies more systematically demonstrated that the bi-parental analysis could also be extended by using genome-wide association study (*GWAS*) based on the linkage disequilibrium mapping, in which association between genotypes and phenotypes is scrutinized over a large germplasm collection (e.g., Remington et al. 2001). This development in association genetics may enhance opportunities for gene-based physiological modelling, especially with development of genome-wide surveys of variation using high-throughput genotyping tools (such as SNP – single-nucleotide polymorphism) across crop germplasm collections (McNally et al. 2009; Huang et al. 2010; Jahn et al. 2011). This development in association genetics may enhance opportunities for gene-based crop modelling. So-called gene-based modelling has empirically been practised by White and Hoogenboom (1996), Messina et al. (2006), White et al. (2008), Zheng et al. (2013), and Bogard et al. (2014), who predicted

flowering and yield traits of crop cultivars via regressing input parameters against binary values of a few relevant candidate genes or markers (also see Chap. 8 of this book by Boote et al.). The SNP-based GWAS can detect many genes and unravel their functions, thereby enabling model-input parameters to be potentially related to many genes. Research on GWAS-based crop modelling is now in the pipeline. However, such an analysis requires considerable attention to population structure and size to satisfy its required statistical power (Hamblin et al. 2011).

9.6 Concluding Remarks

Crop physiology research, traditionally working typically on several distinctive genotypes, has not contributed much to breeding so far. Furthermore, there has been an imbalance in gaining insight and data as geneticists and physiologists seemed to do the things separately. However, the evidence reviewed in this chapter suggests that there is now an opportunity to do a better job because we have more knowledge about QTL (or gene)-function relationships and we have better analytical tools such as QTL (or gene)-based models, dealing with relationships in a genetic population. An integrated QTL-based modelling also provides a common platform for physiologists and geneticists of working all in a synchronous and balanced way, thereby being much more effective in terms of resource use and synergy between approaches. Growing studies on functional genomics and molecular biology will increasingly enable the elucidation of the molecular genetic basis of agronomically and physiologically relevant traits for crop improvement. In the meantime, high-throughput facilities to phenotype a large population for various crop traits, sometimes with high-resolution, are increasingly becoming available. Future crop models should face this unprecedented opportunity. On one hand, model-input parameters should be designed either to be close to those traits breeders, geneticists and biologists commonly score or to be easily measured by modern high-throughput phenotyping facilities, as the optimisation procedure that current crop models often rely on to estimates their parameters (see Chap. 5 of this book by Luquet et al.) may involve another round of uncertainty. On the other hand, model structure and algorithms have to be upgraded, which has been stressed in various preceding chapters of this book and will be further elaborated in the next chapter.

References

Adachi S, Baptista L, Sueyoshi T, Murata K, Yamamoto T, Ebitani T, Ookawa T, Hirasawa T (2014) Introgression of two chromosome regions for leaf photosynthesis from an *indica* rice into background of a *japonica* rice. J Exp Bot 65:2049–2056

Adachi S, Nakae T, Uchida M, Soda K, Takai T, Oi T, Yamamoto T, Ookawa T, Miyake H, Yano M, Hirasawa T (2013) The mesophyll anatomy enhancing CO_2 diffusion is a key trait for improving rice photosynthesis. J Exp Bot 64:1061–1072

Aggarwal PK, Kropff MJ, Cassman KG, ten Berge HFM (1997) Simulating genotypic strategies for increasing rice yield potential in irrigated, tropical environments. Field Crops Res 51:5–17

Amelong A, Gambin B, Severini AD, Borrás L (2015) Predicting maize kernel number using QTL information. Field Crops Res 172:119–131

Baenziger PS, McMaster GS, Wilhelm WW, Weiss A, Hays CJ (2004) Putting genes into genetic coefficients. Field Crops Res 90:133–143

Bertin N, Martre P, Génard M, Quilot B, Salon C (2010) Under what circumstances can process-based simulation models link genotype to phenotype for complex traits? Case-study of fruit and grain quality traits. J Exp Bot 61:955–967

Bogard M, Ravel C, Paux E, Bordes J, Balfourier F, Chapman SC, Le Gouis J, Allard V (2014) Predictions of heading date in bread wheat (*Triticum aestivum* L.) using QTL-based parameters of an ecophysiological model. J Exp Bot 65:5849–5865

Boote KJ, Tollenaar M (1994) Modeling genetic yield potential. In: Boote KJ et al (eds) Physiology and determination of crop yield. ASA/CSSA/SSSA, Madison, pp 533–565

Boote KJ, Kropff MJ, Bindraban PS (2001) Physiology and modelling of traits in plants: implications for genetic improvement. Agric Syst 70:395–420

Buck-Sorlin GH (2002) The search for QTL in barley (*Hordeum vulgare* L.) using a new mapping population. Cell Mol Biol Lett 7:523–535

Chapman SC, Cooper M, Podlich D, Hammer GL (2003) Evaluating plant breeding strategies by simulating gene action and dryland environment effects. Agron J 95:99–113

Chenu K, Chapman SC, Tardieu F, McLean G, Welcker C, Hammer GL (2009) Simulating the yield impacts of organ-level quantitative trait loci associated with drought response in maize: a "gene-to-phenotype" modeling approach. Genetics 183:1507–1523

Condon AG, Richards RA, Rebetzke GJ, Farquhar GD (2004) Breeding for high water-use efficiency. J Exp Bot 55:2447–2460

Dingkuhn M, de Vries Penning FWT, Miezan KM (1993) Improvement of rice plant type concepts: systems research enables interaction of physiology and breeding. In: de Vries Penning FWT et al (eds) Systems approaches for agricultural development. Kluwer, Dordrecht, pp 19–35

Dua AB, de Vries Penning FWT, Seshu DV (1990) Simulation to support evaluation of the production potential of rice varieties in tropical climates. Trans Am Soc Agric Eng 33:1185–1194

Dwivedi SL, Crouch JH, Mackill DJ, Xu Y, Blair MW, Ragot M, Upadhyaya HD, Ortiz R (2007) The molecularization of public sector crop breeding: progress, problems and prospects. Adv Agron 95:163–318

Evans LT, Dunstone RL (1970) Some physiological aspects of evolution in wheat. Aust J Biol Sci 23:725–742

Gu J, Yin X, Struik PC, Stomph TJ, Wang H (2012a) Using chromosome introgression lines to map quantitative trait loci for photosynthesis parameters in rice (*Oryza sativa* L.) leaves under drought and well watered field conditions. J Exp Bot 63:455–469

Gu J, Yin X, Stomph TJ, Wang H, Struik PC (2012b) Physiological basis of genetic variation in leaf photosynthesis among rice (*Oryza sativa* L.) introgression lines under drought and well-watered conditions. J Exp Bot 63:5137–5153

Gu J, Yin X, Stomph TJ, Struik PC (2014a) Can exploiting natural genetic variation in leaf photosynthesis contribute to increasing rice productivity? A simulation analysis. Plant Cell Environ 37:22–34

Gu J, Yin X, Zhang C, Wang H, Struik PC (2014b) Linking ecophysiological modelling with quantitative genetics to support marker-assisted crop design for improved rice (*Oryza sativa* L.) yields under drought stress. Ann Bot 114:499–511

Hamblin MT, Buckler ES, Jannink J-L (2011) Population genetics of genomics-based crop improvement methods. Trends Plant Sci 27:98–106

Hammer G, Cooper M, Tardieu F, Welch S, Walch B, van Eeuwijk F, Chapman S, Podlich D (2006) Models for navigating biological complexity in breeding improved crop plants. Trends Plant Sci 11:587–593

Haverkort AJ, Kooman PL (1997) The use of systems analysis and modeling of growth and development in potato ideotyping under conditions affecting yields. Euphytica 94:191–200

Haverkort AJ, Struik PC, Visser RGF, Jacobsen E (2009) Applied biotechnology to combat late blight in potato caused by *Phytophthora infestans*. Potato Res 52:249–264

Heuvelink E, Marcelis LFM, Bakker MJ, van der Ploeg A (2007) Use of crop growth models to evaluate physiological traits in genotypes of horticultural crops. In: Spiertz JHJ, Struik PC, van Laar HH (eds) Scale and complexity in plant systems research: gene-plant-crop relations. Springer-Verlag, Dordrecht, pp 223–233

Huang N, Angeles E, Domingo J, Magpantay G, Singh S, Zhang G, Kumaravadivel N, Bennett J, Khush G (1997) Pyramiding of bacterial blight resistance genes in rice: marker-assisted selection using RFLP and PCR. Theor Appl Genet 95:313–320

Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Li W, Lin Z, Buckler ES, Qian Q, Zhang QF, Li J, Han B (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42:961–967

Huang X, Paulo MJ, Boer M, Effgen S, Keizer P, Koornneef M, van Eeuwijk FA (2011) Analysis of natural allelic variation in Arabidopsis using a multiparent recombinant inbred line population. Proc Natl Acad Sci U S A 108:4488–4493

Jackson P, Robertson M, Cooper M, Hammer G (1996) The role of physiological understanding in plant breeding, from a breeding perspective. Field Crops Res 49:11–37

Jahan E, Amthor JS, Farquhar GD, Trethowan R, Barbour MM (2014) Variation in mesophyll conductance among Australian wheat genotypes. Funct Plant Biol 41:568–580

Jahn CE, Mckay JK, Mauleon R, Stephens J, McNally KL, Bush DR, Leung H, Leach JE (2011) Genetic variation in biomass traits among 20 diverse rice varieties. Plant Physiol 155:157–168

Jena K, Mackill D (2008) Molecular markers and their use in marker-assisted selection in rice. Crop Sci 48:1266–1276

Kearsey M, Farquhar AGL (1998) QTL analysis in plants; where are we now? Heredity 80:137–142

Kearsey MJ, Pooni HS (1996) The genetical analysis of quantitative traits. Chapman & Hall, London

Khan MS (2012) Assessing genetic variation in growth and development of potato. Ph.D. thesis, Wageningen University, The Netherlands

Khan MS, Yin X, van der Putten PEL, Struik PC (2014) An ecophysiological model analysis of yield differences within a set of contrasting cultivars and an F1 segregating population of potato (*Solanum tuberosum* L.) grown under diverse environments. Ecol Model 290:146–154

Kindu GA, Tang J, Yin X, Struik PC (2014) Quantitative trait locus analysis of nitrogen use efficiency in barley (*Hordeum vulgare* L.). Euphytica 199:207–221

Koornneef M, Stam P (2001) Changing paradigms in plant breeding. Plant Physiol 125:156–159

Kropff MJ, Haverkort AJ, Aggarwal PK, Kooman PL (1995) Using systems approaches to design and evaluate ideotypes for specific environments. In: Bouma J et al. (eds) Eco-regional approaches for sustainable land use and food production. Kluwer, Dordrecht, pp 417–435

Langridge P, Fleury D (2011) Making the most of 'omics' for crop breeding. Trends Plant Sci 29:33–40

Laperche A, Devienne-Baret F, Maury O, Le Gouis J, Ney B (2006) A simplified conceptual model of carbon/nitrogen functioning for QTL analysis of winter wheat adaptation to nitrogen deficiency. Theor Appl Genet 113:1131–1146

Lauteri M, Haworth M, Serraj R, Monteverdi MC, Centritto M (2014) Photosynthetic diffusional constraints affect yield in drought stressed rice cultivars during flowering. PLoS One 9(10):e109054

Lee M (1995) DNA markers and plant breeding programs. Adv Agron 55:265–344

Loomis RS, Rabbinge R, Ng E (1979) Explanatory models in crop physiology. Annu Rev Plant Physiol 30:339–367

Mavromatis T, Boote KJ, Jones JW, Irmak A, Shinde D, Hoogenboom G (2001) Developing genetic coefficients for crop simulation models with data from crop performance trials. Crop Sci 41:40–51

McNally KL, Childs KL, Bohnert R et al (2009) Genome wide SNP variation reveals relationships among landraces and modern varieties of rice. Proc Natl Acad Sci U S A 106:12273–12278

Messina CD, Jones JW, Boote KJ, Vallejos CE (2006) A gene-based model to simulate soybean development and yield responses to environment. Crop Sci 46:456–466

Messina CD, Podlich D, Dong Z, Samples M, Cooper M (2011) Yield-trait performance landscapes: from theory to application in breeding maize for drought tolerance. J Exp Bot 62:855–868

Nakagawa H, Yamagishi J, Miyamoto N, Motoyamo M, Yano M, Nemoto K (2005) Flowering response of rice to photoperiod and temperature: a QTL analysis using a phenological model. Theor Appl Genet 110:778–786

Ohsumi A, Takai T, Ida M, Yamamoto T, Arai-Sanoh Y, Yano M, Ando T, Kondo M (2011) Evaluation of yield performance in rice near-isogenic lines with increased spikelet number. Field Crops Res 120:68–75

Parent B, Tardieu F (2014) Can current crop models be used in the phenotyping era for predicting the genetic variability of yield of plants subjected to drought or high temperature? J Exp Bot 65:6179–6189

Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD (1988) Resolution of quantitative factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721–726

Penning de Vries FWT (1991) Improving yields: designing and testing VHYVs. In: FWT Penning de Vries et al. (eds) Systems simulations at IRRI. IRRI research paper 151, IRRI, Los Baños, pp 13–19

Prioul JL, Quarrie S, Causse M, de Vienne D (1997) Dissecting complex physiological functions through the use of molecular quantitative genetics. J Exp Bot 48:1151–1163

Prudent M, Lecomte A, Bouchet J-P, Bertin N, Causse M, Génard M (2011) Combining ecophysiological modelling and quantitative trait locus analysis to identify key elementary processes underlying tomato fruit sugar concentration. J Exp Bot 62:907–919

Quilot B, Génard M, Kervella J, Lescourret F (2004) Analysis of genotypic variation in fruit flesh total sugar content via an ecophysiological model applied to peach. Theor Appl Genet 109:440–449

Quilot B, Kervella J, Génard M, Lescourret F (2005) Analysing the genetic control of peach fruit quality through an ecophysiological model combined with a QTL approach. J Exp Bot 56:3083–3092

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES IV (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci U S A 98:11479–11484

Reuning GA, Bauerle WL, Mullen JL, Mckay JK (2015). Combining quantitative trait loci analysis with physiological models to predict genotype-specific transpiration rates. Plant Cell Environ 38:710–717

Reymond M, Muller B, Leonardi A, Charcosset A, Tardieu F (2003) Combining quantitative trait loci analysis and an ecophysiological model to analyze the genetic variability of the responses of maize leaf growth to temperature and water deficit. Plant Physiol 131:664–675

Semenov MA, Halford NG (2009) Identifying target traits and molecular mechanisms for wheat breeding under a changing climate. J Exp Bot 60:2791–2804

Setter TL, Conocono EA, Egdane JA, Kropff MJ (1995) Possibility of increasing yield potential of rice by reducing panicle height in the canopy. I. Effects of panicles on light interception and canopy photosynthesis. Aust J Plant Physiol 22:441–451

Shorter R, Lawn RJ, Hammer GL (1991) Improving genotypic adaptation in crops – a role for breeders, physiologists and modelers. Exp Agric 27:155–175

Spitters CJT, Schapendonk AHCM (1990) Evaluation of breeding strategies for drought tolerance in potato by means of crop growth simulation. Plant and Soil 123:193–203

Stam P (1998) Crop physiology, QTL analysis and plant breeding. In: Lambers H et al. (eds) Inherent variation in plant growth: physiological mechanisms and ecological consequences. Backhuys Publishers, Leiden, pp 429–440

Stratton DA (1998) Reaction norm functions and QTL-environment interactions for flowering time in *Arabidopsis thaliana*. Heredity 81:144–155

Struik PC (2010) Can crop physiology help to combat late blight? Potato Res 53(4):277–287

Tardieu F (2003) Virtual plants: modelling as a tool for the genomics of tolerance to water deficit. Trends Plant Sci 8:9–14

Taylaran RD, Adachi S, Ookawa T, Usuda H, Hirasawa T (2011) Hydraulic conductance as well as nitrogen accumulation plays a role in the higher rate of leaf photosynthesis of the most productive variety of rice in Japan. J Exp Bot 62:4067–4077

Teng S, Qian Q, Zeng D, Kunihiro Y, Fujimoto K, Huang D, Zhu L (2004) QTL analysis of leaf photosynthetic rate and related physiological traits in rice (*Oryza sativa* L.). Euphytica 135:1–7

Tuberosa R, Salvi S (2006) Genomics-based approaches to improve drought tolerance of crops. Trends Plant Sci 11:405–411

Uptmoor R, Schrag T, Stützel H, Esch E (2008) Crop model based QTL analysis across environments and QTL based estimation of time to floral induction and flowering in *Brassica oleracea*. Mol Breed 21:205–216

Uptmoor R, Li J, Schrag T, Stützel H (2012) Prediction of flowering time in *Brassica oleracea* using a quantitative trait loci-based phenology model. Plant Biol 14:179–189

Vales MI, Schon CC, Capettini F, Chen XM, Corey AE, Mather DE, Mundt CC, Richardson KL, Sandoval-Islas JS, Utz HF, Hayes PM (2005) Effect of population size on the estimation of QTL: a test using resistance to barley stripe rust. Theor Appl Genet 111:1260–1270

van Eeuwijk FA, Malosetti M, Yin X, Struik PC, Stam P (2005) Statistical models for genotype by environment data: from conventional ANOVA models to eco-physiological QTL models. Aust J Agric Res 56:883–894

Virk PS, Ford-lloyd BV, Jackson MT, Pooni HS, Clemeno TP, Newbury HJ (1996) Predicting quantitative variation within rice germplasm using molecular markers. Heredity 76:296–304

Visker MHPW, Keizer LCP, van Eck HJ, Jacobsen E, Colon LT, Struik PC (2003) Can the QTL for late blight resistance on potato chromosome 5 be attributed to foliage maturity type? Theor Appl Genet 106:317–325

Welcker C, Sadok W, Dignat G, Renault M, Salvi S, Charcosset A, Tardieu F (2011) A common genetic determinism for sensitivities to soil water deficit and evaporative demand: meta-analysis of quantitative trait loci and introgression lines of maize. Plant Physiol 157:718–729

White JW (2006) From genome to wheat: emerging opportunities for modelling wheat growth and development. Eur J Agron 25:79–88

White JW, Hoogenboom G (1996) Simulating effects of genes for physiological traits in a process-oriented crop model. Agron J 88:416–422

White JW, Herndl M, Hunt LA, Payne TS, Hoogenboom G (2008) Simulation-based analysis of effects of *Vrn* and *Ppd* loci on flowering in wheat. Crop Sci 48:678–687

Xu L, Henke M, Zhu J, Kurth W, Buck-Sorlin GH (2011) A functional–structural model of rice linking quantitative genetic information with morphological development and physiological processes. Ann Bot 107:817–828

Yin X, Chasalow S, Dourleijn CJ, Stam P, Kropff MJ (2000a) Coupling estimated effects of QTLs for physiological traits to a crop growth model: predicting yield variation among recombinant inbred lines in barley. Heredity 85:539–549

Yin X, Chasalow S, Stam P, Kropff MJ, Dourleijn CJ, Bos I, Bindraban PS (2002) Use of component analysis in QTL mapping of complex crop traits: a case study on yield in barley. Plant Breed 121:314–319

Yin X, Kropff MJ, Aggarwal PK, Peng S, Horie T (1997) Optimal preflowering phenology of irrigated rice for high yield potential in three Asian environments: a simulation study. Field Crops Res 51:19–27

Yin X, Kropff MJ, Goudriaan J, Stam P (2000b) A model analysis of yield differences among recombinant inbred lines in barley. Agron J 92:114–120

Yin X, Kropff MJ, Stam P (1999a) The role of ecophysiological models in QTL analysis: the example of specific leaf area in barley. Heredity 82:415–421

Yin X, Stam P, Dourleijn CJ, Kropff MJ (1999b) AFLP mapping of quantitative trait loci for yield-determining physiological characters in spring barley. Theor Appl Genet 99:244–253

Yin X, Struik PC (2010) Modelling the crop: from system dynamics to systems biology. J Exp Bot 61:2171–2183

Yin X, Struik PC, Kropff MJ (2004) Role of crop physiology in predicting gene-to-phenotype relationships. Trends Plant Sci 9:426–432

Yin X, Struik PC, van Eeuwijk FA, Stam P, Tang J (2005) QTL analysis and QTL-based prediction of flowering phenology in recombinant inbred lines of barley. J Exp Bot 56:967–976

Zhao XQ, Xu JL, Zhao M, Lafitte R, Zhu L, Fu BY, Gao YM, Li ZK (2008) QTLs affecting morph-physiological traits related to drought tolerance detected in overlapping introgression lines of rice (*Oryza sativa* L.). Plant Sci 174:618–625

Zheng B, Biddulph B, Li D, Kuchel H, Chapman S (2013) Quantification of the effects of *VRN1* and *Ppd-D1* to predict spring wheat (*Triticum aestivum*) heading time across diverse environment. J Exp Bot 64:3747–3761

Zhu XG, Zhang GL, Tholen D, Wang Y, Xin CP, Song QF (2011) The next generation models for crops and agro-ecosystems. Sci China Inf Sci 54:589–597

Chapter 10
Crop Systems Biology: Where Are We and Where to Go?

Xinyou Yin and Paul C. Struik

Abstract The preceding chapters outline approaches in systems biology, genetic mapping and crop modelling, and have shown whether and how these approaches could potentially be integrated to form an effective 'crop systems biology' approach in support of crop improvement. To fulfil the great expectations from the integrated modelling, crop models should be improved based on understandings at lower organizational levels, in the meanwhile ensuring that model-input parameters can be easily phenotyped. The '*crop systems biology*' approach is believed ultimately to realize the expected roles of modelling in narrowing genotype-phenotype gaps and predicting the phenotype from genomic data. Such an approach could be an important tool to solve some imminent food-, feed-, and energy-related, 'real-world' problems.

10.1 Why Crop Systems Biology?

Ecophysiological crop modelling has gradually become a research method and discipline since the first plant modelling paper was published by de Wit (1959). For understanding of a system, models appear good tools for heuristics, for example, to make explicit the importance of properties of system elements in the context of the whole system. For applications, the modelling approach has been predominantly devoted to higher aggregation levels (e.g., optimising agronomic management actions, assessing the impact of climate change on agroecosystems, designing sustaining cropping and farming systems, analysing global yield gaps, etc.). Modelling applications at lower aggregation levels such as designing crop ideotypes and cultivars based on analysing genotype-phenotype relationships have progressed slowly (Jackson et al. 1996; Boote et al. 2013), although use of models as a tool to design crop ideotypes has long been recognised (see review of Loomis et al. 1979).

X. Yin (✉) • P.C. Struik
Department of Plant Sciences, Centre for Crop Systems Analysis, Wageningen University, P.O. Box 430, 6700 AK Wageningen, The Netherlands
e-mail: Xinyou.Yin@wur.nl

X. Yin, P.C. Struik (eds.), *Crop Systems Biology*,
DOI 10.1007/978-3-319-20562-5_10

However, applications for higher and lower aggregation levels cannot be separated absolutely. For example, there have been calls for more mechanistic models to estimate the impact of global CO_2 fertilisation (Yin 2013; Sun et al. 2014). Also, breeding for better crops is essential to improve global food security in the face of climate change. To better address the issue at the higher aggregation levels, developing modelling approaches to study genotype-phenotype relationships becomes increasingly important (Chenu et al. 2009). This latter issue has been addressed internationally either in conference sessions (e.g., Weiss 2003; Cooper and Hammer 2005) or in special symposia or workshops (e.g., Spiertz et al. 2007).

Despite the great effort that has been made, current crop models are mostly crude, in terms of their ability of treating gene-trait-crop relationships. Systems simulation modelling has long been suggested as a powerful tool to understand crop yield formation and to support crop improvement (Loomis et al. 1979). Expectations for modelling in support of modern breeding are high (Dwivedi et al. 2007). However, according to Lawlor (2002), the lack of truly 'mechanistic' crop simulation models (which make use of biochemical information) is a major constraint to advance the understanding of crop yield traits. Boote et al. (2013) also emphasised the needs for more mechanisms in crop models when used for characterising genotype-phenotype relationships. Such a need has been underlined in several chapters of this book.

The modelling studies at the crop level using some knowledge of fundamental plant biology (e.g., biochemistry) are currently sporadic, modelling results published so far to analyse yield traits are inconsistent, and some models are based on untested hypotheses. We, therefore, have proposed a more systematic modelling approach – 'crop systems biology' (Yin and Struik 2007, 2008, 2010), to analyse complex traits at the crop level, not only with the aim of establishing close links with understanding at the gene or genome level, but also in terms of its comprehensive reliance on the whole-metabolism biochemistry and physiology. Therefore, the proposed crop systems biology is a crop-level approach to modelling complex crop traits relevant to global food production and energy supply, via establishing the links between 'omics'-level information, underlying biochemical understanding, and physiological component processes. Crop systems biology, as a research realm, has both fundamental and applied features. For fundamental aspects, crop systems biology models should provide biological interpretation of those phenomena such as genotype-by-environment (G×E) interactions, epistasis, and pleiotropy that prove recalcitrant in genetics. In terms of applications, the goal of crop systems biology is to become a robust tool in support of crop improvement programmes (Yin and Struik 2008).

10.2 Roadmap to Crop Systems Biology; Where Are We?

Development of crop systems biology models certainly depends on what trait a researcher wants to target. Although other traits have also been modelled (e.g., Martre et al. 2003), crop yield is a complex trait that most existing crop simulation

models have attempted to predict. It may not be surprising that simulation of yield formation should be a first major focus trait in crop systems biology. In addition, if crop yield can be modelled accurately, underlying mechanisms for traits related to resource use efficiencies (such as water use efficiency or nitrogen use efficiency) can be analysed accordingly.

This book collected papers from several leading groups, where some initial ideas to develop crop systems biology models have been explored or examples to apply such models in analysing topical crop physiological and breeding questions typically for manipulating crop yield or related traits were described. As introductory material, Chap. 1 by Baldazzi et al. provides some fundamentals of quantitative genetics (particularly mapping of quantitative trait loci QTL), approaches in systems biology for modelling cellular, gene regulatory and metabolic networks, and challenges in integrating these networks into plant or crop models. Chapter 2 by Xu and Buck-Sorlin describes a new, morphologically explicit modelling approach called FSPM (Functional-Structural Plant Modelling) and its potential applications in breeding. Some crop modellers argue that detailed morphological properties have also been captured in classical crop models. Nevertheless, probably because FSPM can create virtual plants visualised in dynamic 3-D pictures, there has been a high level of enthusiasm for applying this FSPM approach, from students to professors, in various aspects of research and education in plant and crop science. Xu and Buck-Sorlin describe how FSPM was linked with QTL analysis, which would assist to breed for various plant traits (morphological and architectural traits in particular). Both systems biology and crop modelling rely on bioinformatics and biometrics or statistics in analysing and interpreting either measured or simulated data. Chapter 3 by Bustos et al. provides statistical approaches (linear mixed models as the default model class) in the context of G×E interactions. While quantitative genetics developed from statistical approaches is fundamental for guiding classical breeding, the factorial regression as discussed in Chap. 3 is a statistical approach closest to the concept of crop growth modelling in capturing the response of physiological and agronomic traits in response to environmental variables. As discussed by Bustos et al., statistical models and crop growth models complement each other. For example, the factorial regression cannot generate tempo-spatial profiles of the trait under study and its associated components. In contrast, crop model simulated responses can be analysed in the context of adapting the crop to the changing environment, allowing the virtual profiling of plants and analysis of how processes interact when crops are perturbed by one or several changes. Chapter 4 by Génard et al. showed how this knowledge generated through *in silico* profiling can be used to decipher G×E interactions so as to build genotypes adapted to particular conditions, i.e., plant ideotypes. Similar line of reasoning is continued in Chap. 5, where Luquet et al. attempted an *in silico* prediction of margins for genetic improvement of rice using Ecomeristem, a model that seems to lie in between FSPM and classical crop models. The target was to analyse the trade-off between early vigour and drought tolerance, and to design rice ideotypes that combine the two traits. One of the traits associated with drought response is the restricted transpiration. In Chap. 6, Sinclair et al. described the steps of modelling-physiology-transcriptomics-genetic screening they followed in developing soybean cultivars with the desired trait that restricts

transpiration. Also based on this restricted transpiration and other examples, Hammer al. argued in Chap. 7 that crop ecophysiology and functional modelling can provide an effective link between molecular and organism scales and can enhance molecular breeding by adding value to genetic prediction approaches. A physiological framework that facilitates dissection and modelling of complex traits can inform phenotyping methods for marker/gene detection and underpin prediction of likely phenotypic consequences of trait and genetic variation in target environments. This is further consolidated in Chap. 8 by Boote et al., who showed model-based approaches revealing that manipulating trait values is beneficial in one environment but not in another environment. They also showed how to link model input parameters with allelic effects of several known genes to establish gene-based modelling of growth and seed yield in common bean, based on the framework of White and Hoogenboom (1996). Perhaps, few model-input parameters are controlled only by pleiotropic effects of a few well characterised major genes; a more likely scenario is that like other quantitative traits, each model-input parameter has own specific polygenes underlying its phenotypic variation. This is a basis of the most active line in this research realm over the last 15 years, i.e., QTL-based crop modelling, and experiences and future prospects are comprehensively reviewed in Chap. 9 by Yin et al.

How far do these states-of-the-art described in the preceding chapters of this book reach the high expectations for crop systems biology? In our judgement, these are just in the juvenile phase of the first of the two-step roadmap that Yin and Struik (2007) proposed for crop systems biology, i.e., the prototype stage and the advanced stage. For the first, a widely used framework or concepts in many existing crop simulation models including processes such as photosynthesis, respiration and assimilate partitioning can still be used. At the level of these processes, there are rich physiological and biochemical data and therefore the understandings are of the highest confidence. For this first step, *crop systems biology* models may not be necessarily more complex than existing crop simulation models in structure, nor is their additional input requirement. The latter is important, and model-input parameters should include those close to the traits breeders score for selection. We should also seek opportunities to derive model-input parameters in parallel with the development of high-throughput phenotyping (White et al. 2012; Parent and Tardieu 2014). However, model algorithms for individual processes are supposed to be more mechanistic than those used in existing crop models. In many cases, a summary form of a detailed biochemical model – e.g., the photosynthesis model of Farquhar et al. (1980) coupled to CO_2 diffusion algorithms (Yin and Struik 2009) – can be incorporated as a sub-model, and this has been incorporated into the crop model GECROS (Yin 2013; Gu et al. 2014). In other cases, direct results or stoichiometries from biochemical studies (e.g., examination of the biochemical pathways for production of proteins, carbohydrates and lipids from glucose by Penning de Vries et al. 1974) can be utilized. A prototype of crop systems biology models needs to be made from this first step, in which physiological and biochemical information at the process level is assembled and then scaled up to the crop level in a way similar to temporal and spatial integrations as practised in conventional crop simulation models. In

relation to crop improvement, a key element of the first step would be to identify the parts of mechanisms that are conservative in energy and water transfer, and in carbon and nitrogen metabolisms, and the parts of mechanisms that show genetic variation and are potentially amenable to selection. In case of grain yield, the prototype models should allow identification and quantitative assessment of specific parts of processes, which could be altered to achieve improvement of yield potential. The parts showing genetic variation can be identified by genetic analysis. For example, in Chap. 7, Hammer et al. indicated that the crop model APSIM has been recently upgraded to structure a generic cereal template for more explanatory approaches to modelling the hierarchy of physiological determinants of crop growth and development. They showcased the stay-green phenotype in sorghum, which was generated as an emergent consequence of canopy nitrogen dynamics associated with genetic differences in dwarfing. Taller genotypes required more nitrogen for structural stem tissue, leaving less available for leaves, which was more rapidly diminished by translocation to grain during grain-filling. Hence, "stay-green" was generated as an emergent consequence in the shorter genotypes in response to genetic differences in plant height.

10.3 Roadmap to Crop Systems Biology; Where to Go?

Perhaps in parallel with the first step, crop systems biology modelling could move to the second step as progresses at the 'omics'-level understanding are being made, towards reaching down to lower organizational levels. For this, it is necessary to map the organization levels and the communication systems between these levels for the different key processes (Struik et al. 2007). Modelling for reaching down to the lower levels is most likely to be done in a manner of one-process-at-a-time; and in this respect, a modular design of the model is important to ensure that changes of a sub-model will not affect other parts of the model. Welch et al. (2003) have already developed a neural network model of *Arabidopsis* flowering time control, based on studies on qualitative, genetic characterization of major flowering time genes in this model plant species. Wilczek et al. (2009) continued the work, using the concept of dynamic simulation as commonly used in crop modelling, by linking individual model coefficients to the activities of specific genes and their regulators involved in the transitions to flowering in *Arabidopsis thaliana*. Similar modelling studies could be performed for phenology of crop species (see an example for maize, Dong et al. 2012, and for wheat, Brown et al. 2013). Further, existing modelling of metabolisms, such as the Benson-Calvin cycle of photosynthesis and the photorespiratory cycle (cf. Giersch 2000) and nitrogen assimilation in relation to the activity of key enzymes (e.g., nitrate reductase and glutamine synthetase), could also be added. With the rapid development of functional genomics in the wake of high-throughput technologies, combined studies of physiological components with gene expression profiles should illustrate the function of genes, biochemical pathways and cellular processes that are affected in a coordinated manner (Stitt and Fernie 2003). Such studies should lay the groundwork for elucidating regulatory networks and causal

linkages among gene products, biochemistry and whole-plant physiology. Summary models for a particular metabolism or process are expected to increasingly become available. Sometimes, models capable of assessing the impact of altered biochemical pathways are not necessarily too much more complex than existing models. For example, von Caemmerer (2013) showed that the models that can simulate the photorespiratory bypass and the bicarbonate pumps (both have been explored as targets to reduce the CO_2 compensation point in C_3 photosynthesis) are only slightly more complex than the standard photosynthesis model of Farquhar et al. (1980). Such summary models can potentially be embedded into crop systems biology models.

Clearly, different temporal, spatial and structural scales are required for different components, pathways, and processes of the model; and this has been showcased in the recent multiscale model of Chew et al. (2014) for *Arabidopsis* that integrates gene dynamics, carbon partitioning, organ architecture, and development response to endogenous and environmental signals. Chapter 1 of this book by Baldazzi et al. discussed the challenges in the multiscale modelling by combining information from molecular biology and genetics with crop models in relation to environmental factors and agricultural practices. Ultimately, crop systems biology may develop into a highly computer-intensive discipline. Such coupled models should inform how and where those recalcitrant genetic phenomena (G×E interactions, epistasis and pleiotropy) come about. They should also allow identification of specific parts of metabolic pathways and processes, which could be altered via genetic engineering to achieve improvement of crop yield potential (Zhu et al. 2011). These specific parts should be amenable to the analysis by the 'omics' approach in terms of the expressions of specific genes, proteins or enzyme activities. For example, gene expression of aquaporins, the putative proteins involved in regulation of water and CO_2 diffusion inside leaves (see also Chap. 6 of this book by Sinclair et al.) have been found to explain most of the variation of stomatal and mesophyll conductance during water stress and recovery in olive (Perez-Martin et al. 2014). In short, these models should ultimately enable *in silico* assessment of crop response to genetic fine-tuning under defined environmental scenarios, thereby being powerful tools in supporting breeding or genetic engineering for complex crop traits. Again, the parsimony rule, especially in terms of the number of required parameters, also applies to models for navigating biological complexity across scales (Hammer et al. 2006), as estimating many parameters in any model is a daunting task (see also Chap. 8 by Boote et al.), even when using advanced bioinformatics or data mining tools (Martin et al. 2015). After all, model sophistication should not be achieved at the cost of model heuristics.

10.4 Crop Systems Biology Needs Cross-Discipline Efforts in Concert

Manipulation of a relatively small number of genes (notably, dwarfing and photoperiod-insensitivity genes in many crops) had resulted in the first 'Green Revolution'. For the next 'Green Revolution' to happen, we have to deal with many

genes so that they work in concert. Advances in genetics and genomics, when integrated via systems biology approaches, can offer unprecedented opportunities to examine crop genetic variation and utilize this variability for breeding purposes in different target environments. However, alterations made at the genome level, although substantial, could have little effect on the crop-level phenotypes (Sinclair et al. 2004; Yin and Struik 2008). Systems biology should not be the privilege of only those working on molecular, sub-cellular or cellular levels. To allow genomics and systems biology to have significant impact, the information from fundamental plant biology should reach up to the crop level, and 'crop systems biology' should be established to deal with complex 'gene-trait-crop' relationships and to enhancing the prediction of the phenotype from genomic information. Recent work of Chew et al. (2014) has shown the promise of such a multiscale approach, based on understandings for the model species *Arabidopsis*. For that to happen for crop species, it is necessary to have the long-term, multi-disciplinary efforts to build the links between geneticists, systems biologists, breeders and crop ecophysiologists towards the next 'Green Revolution' to solve some imminent food-, feed-, and energy-related, 'real-world' problems. By then, the chain as envisaged in Chap. 6 by Sinclair et al., i.e., 'from model to phenotype to genotype to cultivar', can become a reality more than ever. However, at this moment, as expressed in Chap. 8 by Boote et al., "the disciplines have diverged so much that geneticists are not well connected with the field phenotyping, and crop modellers are not connected with the geneticists".

We hope that the publication of this book on crop systems biology promotes cross fertilization between disciplines and can catalyse some joint efforts from the science community to correct that divergence.

References

Boote KJ, Jones JW, White JW, Asseng S, Lisaso JI (2013) Putting mechanisms into crop production models. Plant Cell Environ 36:1658–1672

Brown HE, Jamieson PD, Brooking IR, Moot DJ, Huth NI (2013) Integration of molecular and physiological models to explain time of anthesis in wheat. Ann Bot 112:1683–1703

Chenu K, Chapman SC, Tardieu F, McLean G, Welcker C, Hammer GL (2009) Simulating the yield impacts of organ-level quantitative trait loci associated with drought response in maize: a "gene-to-phenotype" modeling approach. Genetics 183:1507–1523

Chew YH, Wenden B, Flis A, Mengin V, Taylor J, Davey CL, Tindal C, Thomas H, Ougham HJ, de Reffye P, Stitt M, Williams M, Muetzelfeldt R, Halliday KJ, Millar AJ (2014) Multiscale digital *Arabidopsis* predicts individual organ and whole-organism growth. Proc Natl Acad Sci USA 111(39):E4127–E4136

Cooper M, Hammer GL (2005) Complex traits and plant breeding – can we understand the complexities of gene-to-phenotype relationships and use such knowledge to enhance plant breeding outcomes? Aust J Agric Res 56:869–872

de Wit CT (1959) Potential photosynthesis of crop surfaces. Neth J Agric Sci 7:141–149

Dong Z, Danilevskaya O, Abadie T, Messina C, Coles N, Cooper M (2012) A gene regulatory network model for floral transition of the shoot apex in maize and its dynamic modeling. PLoS One 7, e43450

Dwivedi SL, Crouch JH, Mackill DJ, Xu Y, Blair MW, Ragot M, Upadhyaya HD, Ortiz R (2007) The molecularization of public sector crop breeding: progress, problems and prospects. Adv Agron 95:163–318

Farquhar GD, von Caemmerer S, Berry JA (1980) A biochemical model of photosynthetic CO_2 assimilation in leaves of C_3 species. Planta 149:78–90

Giersch C (2000) Mathematical modelling of metabolism. Curr Opin Plant Biol 3:249–253

Gu J, Yin X, Stomph TJ, Struik PC (2014) Can exploiting natural genetic variation in leaf photosynthesis contribute to increasing rice productivity? A simulation analysis. Plant Cell Environ 37:22–34

Hammer G, Cooper M, Tardieu F, Welch S, Walch B, van Eeuwijk F, Chapman S, Podlich D (2006) Models for navigating biological complexity in breeding improved crop plants. Trends Plant Sci 11:587–593

Jackson P, Robertson M, Cooper M, Hammer G (1996) The role of physiological understanding in plant breeding, from a breeding perspective. Field Crops Res 49:11–37

Lawlor DW (2002) Carbon and nitrogen assimilation in relation to yield: mechanisms are the key to understanding production systems. J Exp Bot 53:773–787

Loomis RS, Rabbinge R, Ng E (1979) Explanatory models in crop physiology. Annu Rev Plant Physiol 30:339–367

Martin L, Cook C, Matasci N, Williams J, Bastow R (2015) Data mining with iPlant: a meeting report from the 2013 GARNet workshop, *Data mining with iPlant*. J Exp Bot 66:1–6

Martre P, Porter JR, Jamieson PD, Triboï E (2003) Modeling grain nitrogen accumulation and protein composition to understand the sink/source regulations of nitrogen remobilization for wheat. Plant Physiol 133:1959–1967

Parent B, Tardieu F (2014) Can current crop models be used in the phenotyping era for predicting the genetic variability of yield of plants subjected to drought or high temperature? J Exp Bot 65:6179–6189

Penning de Vries FWT, Brunsting AHM, van Laar HH (1974) Products, requirements and efficiency of biosynthesis: a quantitative approach. J Theor Biol 45:339–377

Perez-Martin A, Michelazzo C, Torres-Ruiz JM, Flexas J, Fernandez JE, Sebastiani L, Diaz-Espejo A (2014) Regulation of photosynthesis and stomatal and mesophyll conductance under water stress and recovery in olive trees: correlation with gene expression of carbonic anhydrase and aquaporins. J Exp Bot 65:3143–3156

Sinclair TR, Purcell LC, Sneller CH (2004) Crop transformation and the challenge to increase yield potential. Trends Plant Sci 9:70–75

Spiertz JHJ, Struik PC, van Laar HH (2007) Scale and complexity in plant systems research: gene-plant-crop relations, vol 21, Wageningen UR frontis series. Springer, Dordrecht, 329 pp

Stitt M, Fernie AR (2003) From measurements of metabolites to metabolomics: an 'on the fly' perspective illustrated by recent studies of carbon–nitrogen interactions. Curr Opin Biotechnol 14:136–144

Struik PC, Cassman KG, Koornneef M (2007) A dialogue on interdisciplinary collaboration to bridge the gap between plant genomics and crop sciences. In: Spiertz JHJ, Struik PC, van Laar HH (eds) Scale and complexity in plant systems research: gene-plant-crop relations. Springer, Dordrecht, pp 319–328

Sun Y, Gu L, Dickinson RE, Norby RJ, Pallardy SG, Hoffman FM (2014) Impact of mesophyll diffusion on estimated global land CO_2 fertilization. Proc Natl Acad Sci U S A 111:15774–15779

von Caemmerer S (2013) Steady-state models of photosynthesis. Plant Cell Environ 36:1617–1630

Weiss A (2003) Introduction. Agron J 95:1–3

Welch SM, Roe JL, Dong Z (2003) Genetic neural network model of flowering time control in *Arabidopsis thaliana*. Agron J 95:71–81

White JW, Hoogenboom G (1996) Simulating effects of genes for physiological traits in a process-oriented crop model. Agron J 88:416–422

White JW, Andrade-Sanchez P, Gore MA, Bronson KF, Coffelt TA, Conley MM, Feldmann KA, French AN, Heun JT, Hunsaker DJ, Jenk MA, Kimball BA, Roth RL, Strand RJ, Thorp KR, Wall GW, Wang G (2012) Field-based phenomics for plant genetics research. Field Crop Res 133:101–112

Wilczek AM, Roe JL, Knapp MC, Cooper MD, Lopez-Gallego C, Martin LJ, Muir CD, Sim S, Walker A, Anderson J, Egan JF, Moyers BT, Petipas R, Giakountis A, Charbi E, Coupland G, Welch SM, Schmitt J (2009) Effects of genetic perturbation on seasonal life history plasticity. Science 323:930–934

Yin X (2013) Improving ecophysiological simulation models to predict the impact of elevated atmospheric CO_2 concentration on crop productivity. Ann Bot 112:465–475

Yin X, Struik PC (2007) Crop systems biology: an approach to connect functional genomics with crop modelling. In: Spiertz JHJ, Struik PC, van Laar HH (eds) Scale and complexity in plant systems research: gene-plant-crop relations. Springer, Dordrecht, pp 61–71

Yin X, Struik PC (2008) Applying modelling experiences from the past to shape crop systems biology: the need to converge crop physiology and functional genomics. New Phytol 179:629–642

Yin X, Struik PC (2009) C_3 and C_4 photosynthesis models: an overview from the perspective of crop modelling. NJAS Wagening J Life Sci 57:27–38

Yin X, Struik PC (2010) Modelling the crop: from system dynamics to systems biology. J Exp Bot 61:2171–2183

Zhu XG, Zhang GL, Tholen D, Wang Y, Xin CP, Song QF (2011) The next generation models for crops and agro-ecosystems. Sci China Inf Sci 54:589–597

Index

X. Yin, P.C. Struik (eds.), *Crop Systems Biology*,
DOI 10.1007/978-3-319-20562-5