

# Chapter 6

## Emerging Concepts and Strategies for Genomics and Breeding

Akshay Talukdar and Pranab Talukdar

**Abstract** To feed the ever-increasing population on earth, production of food crops must increase at an unprecedented pace with limited inputs and little or no harm to the environment. The target is more challenging in the face of a changing climate scenario. The main focus should be on developing technologies and crop genotypes suitable for the input-poor and low-yielding areas that represent the lion's share of the cultivable areas of the world. Plant breeding is evolving; more so with the advancement of molecular biological sciences. Emerging concepts of structural and functional genomics, transcriptomics, proteomics, and metabolomics approaches contribute towards identification of target genes and eQTL (expression-quantitative trait loci) for effective deployment. The genome editing technologies creating site-directed mutation can facilitate development of nontransgenic designer crop genotypes having wider adaptation. Plant breeding approaches should focus on the natural resources to identify and deploy useful gene(s)/alleles to develop genotypes with enhanced yield potential, better stress tolerance and, quality end products. Improvement in mapping efforts like genomic selection (GS), marker-assisted recurrent selection (MARS), and next-generation mapping, viz. NAM (nested association mapping), MAGIC (multiparent advanced generation intercross) would accelerate breeding progress with improved genotyping and phenotyping facilities. Plant biotechnology and genetic engineering techniques would continue to play a pivotal role in making a crop widely adaptable to the changed climate. Improving crop efficiencies in utilization of solar radiation, inorganic nitrogen, water and other inputs would render crops suitable for

---

A. Talukdar (✉)

Division of Genetics, Indian Agricultural Research Institute, New Delhi 110012, India  
e-mail: [atalukdar@iari.res.in](mailto:atalukdar@iari.res.in)

P. Talukdar

Department of Plant Breeding and Genetics, Assam Agricultural University, Jorhat 785013, Assam, India

climate-resilient agriculture. Policies should be in place to make technologies accessible and affordable by all sections of users globally.

## 6.1 Introduction

Crop yield is a complex trait; it depends not only on genetic factors but to a larger extent on the nongenetic factors as well. Therefore, manipulating the yield potential of crops to meet the ever-increasing demand for food is a big challenge. Given current agricultural growth and crop production, it is quite unlikely that the target of feeding 9.1 billion people on the Earth by 2050 will be met easily. It has been predicted that the current production has to be increased by 70 % to meet the target (FAO 2011). However, the approaches adopted to achieve this goal have to be quite different from those used in the past; it cannot be as harsh to the environment as previous approaches. Therefore, the strategies should be to maximize crop production with less inputs and energies, and with the least or no harm to the environment; this is where science in general and, plant breeding, genetics and genomics in particular have a pivotal role to play.

As crop yield is highly influenced by the environment, the focus should be on beating the environment where the crop is grown. Typically, a crop variety with higher yield potential may not grow well and produce satisfactorily if the required inputs and growing conditions are not met. A larger part of the global crop-producing areas are of the poor yielding or low-input type. Specific approaches and strategies are needed to enhance yield in such environments. Crop varieties designed especially for such situations should therefore be the focus of the climate-resilient crop breeding program in the future.

Reduction of yield loss is indirectly yield gain. More attention is needed in tackling yield-reducing factors such as drought, salinity, heat, flood, frost, etc., whose intensities are predicted to be greater in the changing climate scenario. Plant breeding can play an important role in reducing yield loss by developing varieties suited to local stress conditions and by making them more resilient to biotic (e.g., insects, diseases, and viruses) and abiotic challenges (e.g., droughts and floods). Genes conferring increased drought tolerance may also have a widespread impact on yield (Nelson et al. 2007). Studies estimated that the global yield loss due to biotic stresses averages over 23 % of the estimated attainable yield across major cereals (FAOSTAT 2009). Yield stability is the other factor of importance, which in fact is a difficult breeding target. However, there is evidence for genetic control of stability making it an achievable target (Kraakman et al. 2004). Similarly, transgenic approaches are also likely to improve yield stability (Nelson et al. 2007). A single gene was able to substantially increase yield in many crops including rice and wheat leading to the Green Revolution. In parallel with the increase in yield, sustaining it in the changing climatic scenario is another important issue that needs greater emphasis so as to adopt appropriate strategies for development of suitable technologies.

Plant breeding has been evolving; it has transitioned from a skillful “art” to a technology-dependent “science.” Infusion of newer scientific technologies has made plant breeding more precise, productive, and predictable. It has become possible to identify the required plants from segregating progenies even at the seedling stage, representing huge savings in time, energy, and costs. Incorporation of genomics and reverse genetics approaches in plant breeding has shown hope of increasing and sustaining crop production even in changing climatic situations. However, maintaining resistance to rapidly evolving pests and pathogens is going to be an essential mainstay of any breeding program. Interactions between breeders, pathologists, and agronomists must be maintained to ensure that crops and cropping systems change coordinately (Tester and Langridge 2010).

Problems on the agricultural front are emerging with every passing day; challenges are presenting themselves constantly. Newer and more effective strategies and approaches are being used to tackle these problems and challenges. In the field of genetics, genomics, and plant breeding numerous strategies and approaches are being tested regularly. However, only a few are retained over time. In this chapter, emerging strategies and approaches in the field of genomics and plant breeding are discussed in the light of increasing crop production under changing climatic conditions.

## 6.2 Requirement for New Approaches and Strategies in Genomics and Plant Breeding

Agriculture with intense support from science, technology, and regulatory policies has made substantial progress in terms of producing more food worldwide. Global food grain production has increased from 850 million tons (mt) in 1960 to 2,350 mt in 2007 (Godfray et al. 2010). It has saved millions of people from starvation and untimely death globally, in the developing countries in particular. However, the situation now is not rosy, either; it is intensifying and worsening with the passing of time. Further, the challenges faced today are different in terms of their nature, extent, and attributes. The pressure for enhanced food production is enormous owing to the ever-growing population. About 1.0 billion people were added in a span of 12 years from 1999 to 2011 raising the global population from 6.0 to 7.0 billion (Table 6.1). It is predicted that by 2050, there will be 9.1 billion people on earth. To feed such a huge population, our current productivity has to be increased by about 70 % by this time. However, the global cultivated area of over 1.5 billion ha, which represent some 12 % of the world’s land surface, is under tremendous pressure due to rapid urbanization, industrialization, and other nonagricultural uses. As per FAO’s estimate, by 2050, 70 % of the global population will be urban. So, more food has to be produced from limited areas. Secondly, irrigation is crucial for crop production. Agriculture is responsible for about 70 % of all fresh water withdrawn for human uses. Water is becoming a scarcer commodity everyday

**Table 6.1** Growth of world's population

Year	Population <sup>a</sup>	Time span (years)
1	200 mn	
1650	500 mn	1,649
1804	1.0 bn	154
1927	2.0 bn	123
1960	3.0 bn	33
1975	4.0 bn	15
1987	5.0 bn	12
1999	6.0 bn	12
2011	7.0 bn	12
2027 <sup>b</sup>	8.0 bn	16
2050 <sup>b</sup>	9.1 bn	23

Source: FAO (2011)

<sup>a</sup>mn million, bn billion

<sup>b</sup>Projected

leaving less water to produce more food. In large areas in India and China, ground water levels are falling by 1–3 m per year. The quality of produce is becoming a priority among consumers. Energy is becoming costlier leaving less scope not to be efficient. On top of it all, changes in climatic conditions have started showing an impact on crop production through untimely flood, heat waves, and prolonged drought (Ciais et al. 2005; Lobell et al. 2008; Ortiz et al. 2008; Peltonen-Sainio et al. 2010). So, more food has to be produced from a limited area, with less water, minimum energy, and least harm to the environment; a daunting task indeed.

Food security for the growing population is a matter of concern globally. However, the less developed countries are predicted to be particularly vulnerable in terms of food security. The reasons are (1) most of these countries are net importers of cereals (Dixon et al. 2009); (2) many of their national agricultural research services lack sufficient capacity for timely delivery of agricultural technologies (Kosina et al. 2007); and (3) the majority are located in regions that are vulnerable to climate change (Lobell et al. 2008). The net impact of a food crisis would be enormous and would certainly put humanity in threat. It demands preparedness in order to lessen the impact, if one cannot prevent it. The science of genetics, genomics, and plant breeding has to be advanced and utilized fully to meet this challenge.

### 6.3 Genomic Approach

Genomics can be defined as the generation of information about living things by systematic approaches that can be performed on an industrial scale (Brent 2000). The “information” means the genetic message coded in the DNA sequences, the genome of an organism. Such information is revealed through DNA sequencing approaches, which Weinstein (1998) referred to as “omics.” With specialized

objectives and purpose fulfilled, the genomics approaches may be classified as structural, functional, and comparative genomics. Currently, more specialized branches such as proteomics, transcriptomics, and metabolomics have also come into being. All these classes of genomics provide information about genome organization, linkage, protein complement, gene regulation as well as phylogeny and evolutions.

### **6.3.1 Structural Genomics**

One of the specialized branches of genomics is structural genomics, which focuses mainly on the physical aspects of the genome through the construction and comparison of gene maps and sequences, as well as gene discovery, localization, and characterization. With the availability of more and more genome sequences that have been sequenced at a speed never thought of before, it has become imperative to determine structure and assign function to the different genes inscribed in the sequences, a challenge to the scientific community, indeed. However, with advances in the tools of structural biology, the challenge has largely been met through large-scale determination of three-dimensional structural models for all known proteins, protein families or protein domains from which most others can be predicted computationally with a reasonable degree of accuracy. It is believed that the benefit of structural genomics will be much higher than envisaged. The most prominent benefit will be to establish the relationship between one-dimensional sequence information and three-dimensional structure of the protein. Structural genomics may well provide the means of coming to grips with this important intellectual challenge (Burley et al. 1999).

The major goal of structural genomics is to provide a structural template for a large fraction of protein domains. With various approaches, structures of a section of the protein families have been predicted. Approximately 20 % of the known families with three or more members currently have a representative structure. As per estimates, the number of apparent protein families will be considerably larger than previously thought. However, the vast majority of these families will be small, and it will be possible to obtain structural templates for 70–80 % of protein domains with an achievable number of representative structures, by systematically sampling the larger families (Yan and Moult 2005). So, translating the outcome of structural genomics into a practical product that can be applied for human welfare will take some time.

### **6.3.2 Functional Genomics**

The branch of genomics that deals with revealing the function of genes and other parts of the genome is known as functional genomics. It is the emerging field of

molecular biology that is attempting to make use of the vast wealth of data produced by genome sequencing projects to describe genome function in a meaningful way. It primarily uses high-throughput techniques like DNA microarrays, transcriptomics, proteomics, metabolomics, and mutation analysis to describe the function and interactions of genes. The functional genomics approaches are more powerful and robust than the conventional genetic approaches in assessing phenotype largely with respect to the scale and automation involved in the investigation.

Microarray technology, which involves hybridization of unknown samples with a GeneChip containing DNA/RNA probes, is of great use in finding gene/protein function. The use of microarrays for expression profiling was first published in 1995 (Schena et al. 1995) and the first complete eukaryotic genome (*Saccharomyces cerevisiae*) on a microarray was published in 1997 (Lashkari et al. 1997). Plants under biotic and abiotic stresses can be subjected to characterization through microarray analysis. The expression profiles thus generated can provide extensive data for identification of genes involved in the resistance-susceptibility reaction. A first-generation maize GeneChip containing 1,500 ESTs/gene could identify 117 genes that were either induced or repressed 6 h after inoculation with the fungus *Cochliobolus carbonum* (Baldwin et al. 1999). The function of the genes thus identified has to be confirmed on an individual gene basis. A powerful reverse genetics approach like gene-silencing, TILLING, Eco-TILLING, or any other gene knock-out technique should help resolve this issue with better resolution (Marteinssen 1998; Baulcombe 1999).

The variety of methods available for global analysis of protein profiles and cataloging protein-protein interactions on a genome-wide scale are technically complex and demanding. However, newer and improved technology and algorithms for collecting, displaying, and analyzing the vast amounts of quantitative expression data are being developed (Eisen 1998). Such information is yet to be used in plants extensively. However, a lot has been achieved in regard to plant-resistance genes (Michelmore 2000).

### 6.3.3 Comparative Genomics

Comparative genomics is the analysis and comparison of genetic materials from different taxa including species, subspecies and even genera to study evolution, gene function, and other important traits. The purpose is to gain a better understanding of how species have evolved and to determine the function of genes and noncoding regions of the genome. It also provides insight into the uniqueness of and homology between different taxa. Here, comparison of the sequences are made for gene location, gene structure (with respect to exon number, exon length, intron length sequence similarity), and gene characteristics (e.g., splice sites, codon usage, conserved synteny). It involves the use of computer programs that can line up multiple genomes and look for regions of similarity among them. In recent time, a number of species including crop plants have been fully sequenced and, the

sequences are available in the public domain. Such sequences are valuable genomic tools and have been used for gene identification for specific target traits like drought tolerance, thermotolerance and so on.

In the process of evolution, only a finite number of chromosomal rearrangements have occurred in angiosperm plants. Therefore, large blocks of genetic material would be syntenic between the genomes of related species/genera, and this has been well documented in *Arabidopsis* and *Brassica* species (Gale and DeVos 1998). With more and more crop sequences being available, it would be possible to predict the position of most of the genes, if not all, in each part of the genome. However, for prediction of a gene from a sequence, a number of considerations as well as comparisons are required to be made. For simplicity and efficiency of the job of comparison and identifying the number of gene(s) to the maximum correct possible, a number of computer software programs are used. For example, GenScan, GeneWise, Procrustes, Rosseta, SGP1 (Syntetic Gene Prediction), CEM (Conserved Exon Method), GenomeScan, SGP-2, TwinScan, SLAM, and DoubleScan are commonly used for specific purposes. Similarly, a wide range of genome comparison tools is now available, viz., DIALIGN, ASSIRS (Accelerated Search for SImilarity Regions in Chromosomes), MUMmer (Maximal Unique Match (mer)), and GLASS (GLocal Alignment SyStem). Chain et al. (2003) categorized them as (1) pair-wise local alignment comparison tools; (2) global alignment tools: pair-wise alignment, multisequence alignment and multigenome alignment; (3) substrings maximum-exact-match tools; and (4) alignment viewing tools. All of these tools have their own niches, advantages, and limitations; so the user has to determine which one is applicable for a specific purpose.

One of the challenges in comparative genomics is to distinguish “orthologs” from “paralogs,” particularly in large diverse resistance gene families (Michelmore 2000). Similarly, the chromosomal positions of resistance gene candidate sequences seem not to be preserved between grass species (Liestner et al. 1998). For example, homologs of the *RPML* gene are missing from susceptible genotypes of *Arabidopsis* (Grant et al. 1998; Stahl et al. 1999). Further, in several species, resistance genes seem to be either telomeric or close to the centromere. It may happen because chromosome rearrangements often involve changes close to the telomere and centromere; therefore chromosomal position may contribute to the lack of synteny of some resistance genes (Michelmore 2000).

The advantage of sequence similarity is that the PCR primer designed based on such sequences should be applicable for all species having similar sequences. Using such a concept, a large number of resistance-gene candidate sequences have been cloned from diverse species (Meyers et al. 1999; Rivkin et al. 1999; Pan et al. 2000). In addition to gene and regulatory gene identification, comparative genomics may provide information regarding allelic variations, which may be helpful in reasoning the specificity at molecular level. Such study has been reported in resistance genes of crop plants (Ellis et al. 1999).

An important application of comparative genomics has been found in Sorghum (*Sorghum bicolor*) whose genome sequence information has become an important reference for genomics, transcriptomics, and other applications of systems biology

in sugarcane (Paterson et al. 2009; Wang et al. 2010). Phylogenetically, sorghum is very closely related to the *Saccharum* genus and current commercial sugarcane hybrids; both belong to the subtribe Saccharinae of the family Poaceae. Comparison of ESTs from sorghum with sugarcane revealed 97 % mean sequence identities as against 93 % and 86 % with maize and rice, respectively (Paterson et al. 2009). Further, comparison of 20 sugarcane BACs to their respective homologous regions in sorghum led to the identification of 209 and 189 genes from sugarcane and sorghum respectively, based on ESTs and gene-calling algorithms (Paterson et al. 2009; Wang et al. 2010). This shows the practical utility of comparative genomics for accumulating information in less studied crops.

### 6.3.4 Transcriptomics

Transcriptome refers to the complete set of transcripts in a cell, and their quantity, for a particular developmental stage or physiological state. Transcriptomics i.e., the study of the transcriptome aims to comprehensively profile all the information that appears in the RNA pool within a system (e.g., a cell, body fluid, or tissue). Understanding the transcriptome helps in interpreting the functional elements of the genome and revealing the molecular constituents of cells and tissues, and also in understanding development and diseases. The key aims of transcriptomics can be noted as: to catalog all species of transcript, including mRNAs, noncoding RNAs, and small RNAs; to determine the transcriptional structure of genes, in terms of their start sites, 5' and 3' ends, splicing patterns and other posttranscriptional modifications; and to quantify the changing expression levels of each transcript during development and under different conditions (Wang et al. 2009). It is important to note that the information carried in the transcriptome is not necessarily a direct recapitulation of information from the genome, nor is it generated only during DNA-dependent RNA synthesis. In fact, the sequence of an RNA molecule can become modified by a number of processes, including differential splicing and RNA editing. Therefore, transcriptome profiling needs validation through various methods including real-time quantitative polymerase chain reaction (Q-PCR) (Tsiridis and Giannoudis 2006).

A number of technologies varying in approach have been developed for estimation and quantification of the transcriptomes. The hybridization-based approaches typically involve incubating fluorescently labeled cDNA with custom-made microarrays or commercial high-density oligo microarrays. Specialized microarrays have also been designed and are available. Genomic tiling microarrays that represent the genome at high density allow the mapping of transcribed regions to a very high resolution, from several base pairs (bp) to ~100 bp. In general, the hybridization-based approaches are high-throughput and relatively inexpensive. Conversely, the sequence-based approaches directly determine the cDNA sequence. The sequencing of cDNA or EST libraries through Sanger chemistry is relatively low throughput, expensive and generally not quantitative. Tag-based

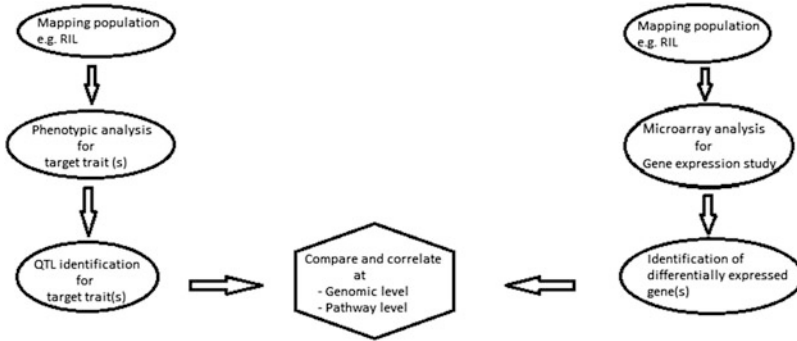


methods including serial analysis of gene expression (SAGE), cap analysis of gene expression (CAGE) and massively parallel signature sequencing (MPSS) are high-throughput and can provide precise, “digital” gene expression levels. In recent time, RNA-Seq (RNA sequencing), a high-throughput DNA sequencing method has been used for both mapping and quantifying transcriptomes. It has successfully been applied to *Scharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Arabidopsis thaliana*, mouse, and human cells (Lister et al. 2008; Nagalakshmi et al. 2008).

For effective identification of candidate gene(s), expression analysis has been combined with genetic or quantitative trait loci (QTL) mapping, an approach called “genetical genomics” (Jansen and Nap 2001). In this approach, total mRNA/cDNA from the tissue/organ of each individual of a mapping population (e.g., recombinant inbred line, RIL) is hybridized with the microarray carrying a high number of cDNA fragments representing the tissue of interest, and quantitative data are recorded against each gene on the filter (de Koning and Haley 2005; Fig. 6.1). The expression data are then subjected to QTL analysis. The QTL so-detected are called expression QTL (eQTL). The genetical genomics unravels genes and gene products that are involved in metabolic and regulatory (e.g., developmental) pathways. For each gene (cDNA) or gene product analyzed in the segregating population, QTL analysis will pinpoint the regions of the genome influencing its expression. Thus genetical genomics can identify candidate gene(s) by combining the QTL information from all genes and gene products that are analyzed. It will indicate what portion of the variation in gene expression maps to the genes themselves (*cis*-acting factors), as opposed to other genomic locations (transacting factors) (Jansen and Nap 2001). The approach of regulatory network construction by combining eQTL and mapping and regulatory candidate gene selection has been used for studying genes associated with flowering behavior in *Arabidopsis* (Keurentjes et al. 2011). With the advent of more sophisticated analytical tools and powerful software for data analysis, it will find more footage in the crop improvement program in the future.

### 6.3.5 *Metabolomics*

Each living cell produces an array of metabolic intermediates or end products valuable for its survival or defense, called metabolites. Levels of metabolites can be related to the response of the cells to the genetic or environmental changes. The set of metabolites synthesized by a biological system constitute its “metabolome” (Oliver et al. 1998). The metabolomes share some basic functional groups viz., hydroxyls, alcohols, steroids, alkyls, benzyl rings, etc. The combinations of these functional groups lead to the development of unique compounds varying in their solubility, stability, melting points, and reactivities typical of plant metabolism (Roessner and Beckles 2009). Simultaneous detection and analysis of such compounds has led to the emergence of a new field of “omics” called “metabolomics,” that is, the description of the metabolic state of a biological



**Fig. 6.1** eQTL combines QTL analysis with gene expression

system in response to environmental and genetic perturbations (Oliver et al. 1998; Fiehn et al. 2001; Villas-Bôas et al. 2007). A number of analytical technologies are needed to enable the separation, detection, and quantification of the metabolomes present in a cell. The most common platforms are liquid and gas chromatography both coupled with mass spectrometry (LC-MS and GC-MS) and nuclear magnetic resonance spectroscopy (NMR).

The primary objective of metabolomics is to associate the relative changes in quantitative metabolite levels with functional assignments so as to understand and predict the behavior of a complex system like plants (Oliver et al. 1998). The metabolic profile provides a readout of the metabolic state of an individual that cannot be obtained directly from DNA genotyping, gene expression, or proteomic profiling analyses. The metabolic changes would be the key in identification of the enzymes involved in the biochemical pathways, which in turn would be linked to the underlying gene. Several studies have been reported using metabolomic approaches for gene function analysis and QTL identification (Kazuki 2006), genotype discrimination (Taylor et al. 2002) as well as metabolite characterization so as to identify the regulatory keys and gene function (Saghatelian et al. 2004). With the advancement of analytical tools and techniques and, sophistication of data handling software, metabolomics holds great promise. The pharmacometabolomic approach (Clayton et al. 2006) is useful for providing information on pathways that contribute to determination of the individual pharmacokinetic and pharmacodynamic behaviors of a drug associated with response as well as insight into the mechanisms responsible for individual variation in terms of drug side effects and toxicity (Corona et al. 2012).

Like other important plant traits such as yield and flowering time, metabolite levels in plant tissues (m-trait) are also a quantitative trait, and QTLs responsible for such m-traits (e.g., level of seed vitamin E) have been identified (Giles 2007). By taking advantage of the high-density linkage map of molecular markers, a number of causal genes responsible for each mQTL could be deduced for further investigation of regulatory systems in complicated plant metabolism pathways. The analysis of the relationship between m-traits and other important agronomic and

biological traits such as yield, taste, and biomass in tomato indicated that there are certain correlations among these traits (Grossman and Takahashi 2001; Schauer et al. 2006). The growth rate of *Arabidopsis* seedlings is to some extent predictable from the metabolome signature (Meyer et al. 2007). In the near future, metabolomics could also play a key role in the evaluation of genetically modified crops, and in understanding plant systems and, developing further biotechnology applications.

### 6.3.6 Proteomics

A central dogma of molecular biology states that the genetic information written in the DNA molecule is passed through mRNA, to ultimately express in terms of protein; the path is not straight forward, though. The presence of posttranslational modification and posttranslational truncation of proteins, and protein–ligand interactions are a few examples that illustrate the complexity of the system. Therefore, the study of proteins has long been important but it gained momentum only during the last decade with the development of technology capable of performing large-scale analyses and identification of proteins (Issaq et al. 2002; Wang and Hanash 2003). This achievement has opened the door for comprehensive studies of proteins related to a genome, called proteomics (Wilkins et al. 1996). Isolation of protein from sample and their identification is more critical. However, introduction of matrix-assisted laser desorption ionization and time-of-flight (MALDI-TOF) mass spectrometry and electro-spray ionization (ESI) tandem mass spectrometry has revolutionized the field.

The technique of proteomics has been used for various purposes including studying the protein–protein interaction and identification of multisubunit complexes. It has already been applied in crops like maize (Chang et al. 2000), chickpea (Bhushan et al. 2007; Pandey et al. 2008), rice, and *Arabidopsis* (Tsugita et al. 1996) to generate useful information for genomics and crop breeding.

## 6.4 Emerging Concepts of Genomics

### 6.4.1 Genome Editing

The genome is the store house of genetic information. So, attaining the ability to read and edit genes in any organism has long been a goal in order to have in-depth knowledge and understanding of the genetic control of any cellular activities, and applying such knowledge to improve agricultural productivity, cure human diseases, and so on. A recently developed genomic technology called “genome editing” has promised to realize this goal through an unmatched level of precision

in studying gene function and biological mechanisms potentially in any system ranging from fruit flies, human cell lines, zebra fish to plants and hosts of other organisms. It enables efficient and precise genetic modification via induction of a double-strand break (DSB) in a specific genomic target sequence, followed by the generation of desired modifications during subsequent DNA break repair (Bibikova et al. 2002, 2003). This DSB is induced by a “zinc finger nuclease” (ZFN) (Kim et al. 1996; Bibikova et al. 2001) or by “transcription activator-like effector (TALE) nuclease (TALEN)” (Wood et al. 2011) at desired loci that can be repaired by the error-prone nonhomologous end-joining (NHEJ) (Jackson and Bartek 2009; Lieber 2010; Moynahan and Jasin 2010) method to yield small insertions and deletions (indels) at the break sites. The ZFN and TALEN can effectively be used to introduce into endogenous loci the targeted modifications, namely “gene disruption,” “gene correction,” and “targeted gene addition.”

**Gene Disruption.** It is the simplest approach of genome editing in which a targeted gene is rendered nonfunctional or knocked out by introduction of error through engineered ZFNs. This process has been used in *Drosophila* (Beumer et al. 2008), mouse (Geurts et al. 2009), and zebrafish (Doyon et al. 2008) with varying degrees of success. In *Drosophila*, ZFNs targeting exonic sequences delivered via mRNA injection into the early fly embryo produced up to 10 % of the progeny with mutation for the gene of interest (Beumer et al. 2008). The success has been found to be more pronounced in zebrafish where up to 50 % germline mosaicism at the targeted gene was reported (Doyon et al. 2008; Meng et al. 2008; Foley et al. 2009). Engineered ZFNs were used to knockout the dihydrofolate reductase (*Dhfr*) gene in Chinese hamster ovary (CHO) cells, a mammal cell line. A plasmid encoding the ZFNs was introduced by transient transfection, which resulted in disruption frequencies of up to 15 % of alleles in the cell population (Santiago et al. 2008).

**Gene Correction.** It involves a homology-directed repair (HDR) mechanism in which a single nucleotide or short heterologous stretches from an episomal donor can be transferred to the chromosome following a ZFN-induced DSB. The endogenous repair machinery uses the investigator-provided donor as a template for repairing the DSB via the synthesis-dependent strand annealing process (Bozas et al. 2009). This technique, also called “allele editing” enables the study of gene function, de novo creation of point mutations at a native locus, and facilitates gene correction. The robustness of this approach was demonstrated in human interleukin-2 receptor- $\gamma$  (*IL2RG*) (Urnov et al. 2005) and three other genes (Maeder et al. 2008).

Gene correction through ZFNs has also been achieved in plants (tobacco) (Townsend et al. 2009); the ZFN-expressing plasmids were codelivered to tobacco protoplasts along with a linear donor molecule encoding a point mutation that corrects the endogenous gene to an herbicide-resistant form. In 75–96 % of all the herbicide-resistant calli, the correction had occurred.

**Gene Addition.** The ZFN driven approach has enabled transfer of gene-sized heterologous DNA sequences from an episomal or linear extrachromosomal donor to the target genome. Initially, this work was optimized in flies; it has now

been extended to human beings where transgenes of up to 8 kb in length were added to the genome through ZFNs in the *IL2RG* gene (Moehle et al. 2007). Similarly, ZFNs were used to generate an isogenic panel of mouse ES cells carrying a defined series of alleles for an endogenous gene (Goldberg et al. 2010).

Good news for the plant biologists is that gene addition through ZFNs has been reported in plants including tobacco and maize. In tobacco, ZFNs targeting an endogenous endochitinase gene successfully added an herbicide-resistance marker in nearly 10 % of the cases (Cai et al. 2009). Similarly in maize, ZFNs targeted to the gene encoding an enzyme required for the production of phytate were introduced with a donor carrying an herbicide-resistance marker (Shukla et al. 2009). The ZFN-edited plants were fertile, and showed normal Mendelian fashion of inheritance for the target trait. The success of this approach has promise for use in “trait stacking” in major crop species like rice, maize, wheat, soybean, etc.

The application of ZFN and TALEN techniques and its potential impact in plant breeding is enormous. Testing candidate loci may become a straightforward task, as locus-specific knockouts or allelic replacement allow both functional validation and a direct means of estimating the effects of individual alleles. It would allow editing alleles at loci of known agronomic interest directly in the individual lines, entirely bypassing the process of backcrossing. A likely possibility of this technique in the near future is the targeted replacement of deleterious mutation in elite breeding lines. Thus, genomic editing is likely to be an attractive and potential alternative to current transgenic technologies (Morrell et al. 2012). However, certain issues like specificities, off-target, process optimization, etc., must be resolved carefully before it can be used routinely.

### 6.4.2 Next-Gen Sequencing

Revealing genetic diversity and putting them to use for crop production is the primary goal of the plant breeding exercises. Modern molecular biological tools including the molecular markers, viz., restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), and simple sequence repeats (SSR), etc., have now been used extensively for this purpose. However, each marker system has its own strengths and limitations. Further, speed, efficiency, and accuracy are the other factors limiting the life and utility of these systems. The single nucleotide polymorphism (SNP) is the sequence-based marker system that had overcome many such limitations. However, the cost involved in sequencing the gene or target sequence in order to generate the SNP is very high and, may even be prohibitive in certain cases.

Since the early 1990s, DNA sequence production has almost exclusively been carried out with capillary-based, semiautomated implementations of Sanger biochemistry (Sanger et al. 1977; Hunkapiller et al. 1991). It takes place in a “cycle sequencing” reaction, in which cycles of template denaturation, primer annealing, and primer extension are performed. Each round of primer extension is

stochastically terminated by the incorporation of fluorescently labeled dideoxynucleotides (ddNTPs). In the resulting mixture of end-labeled extension products, the label on the terminating ddNTP of any given fragment corresponds to the nucleotide identity of its terminal position. Sequence is determined by high-resolution electrophoretic separation of the single-stranded, end-labeled extension products in a capillary-based polymer gel. Laser excitation of fluorescent labels as fragments of discrete lengths exit the capillary provides the readout that is represented in a Sanger sequencing “trace.” Software translates these traces into DNA sequence, while also generating error probabilities for each base-call (Ewing and Green 1998; Ewing et al. 1998). After three decades of gradual improvement, the Sanger biochemistry can be applied to achieve read-lengths of up to ~1,000 bp, and per-base “raw” accuracies as high as 99.999 % with costs on the order of \$0.50 per kilobase (Shendure and Ji 2008).

In order to incorporate speed, accuracy, and automation, several newer and improved technologies have been developed that use cyclic-array sequencing and are generally categorized as next-generation sequencing (NGS) (Shendure and Ji 2008). Basically, the concept of cyclic-array sequencing involves sequencing of a dense array of DNA features by iterative cycles of enzymatic manipulation and imaging-based data collection (Mitra and Church 1999). It has been commercially utilized in a number of products like 454-sequencing, solexa technology, the SOLiD platform, the Polonator, and the HeliScope Single Molecule Sequencer technology.

These platforms have a lot in common and differences in their work flows (Table 6.2). Library preparation is accomplished by random fragmentation of DNA, followed by *in vitro* ligation of common adaptor sequences. The clonally clustered amplicons to serve as sequencing features can be generated by approaches like *in situ* polonies (Mitra and Church 1999), emulsion PCR (Dressman et al. 2003) or bridge PCR (Fedurco et al. 2006). The sequencing is done through synthesis, i.e., serial extension of primed templates through polymerase or ligase (Brenner et al. 2000; Mitra et al. 2003; Shendure et al. 2005; Turcatti et al. 2008). Data are acquired by imaging of the full array at each cycle.

Advantages of second-generation or cyclic-array strategies, relative to Sanger sequencing, includes (1) *in vitro* construction of a sequencing library, followed by *in vitro* clonal amplification to generate sequencing features, which circumvents several bottlenecks that restrict the parallelism of conventional sequencing (that is, transformation of *E. coli* and colony picking). (2) Array-based sequencing enables a much higher degree of parallelism than conventional capillary-based sequencing. As the effective size of sequencing features can be of the order of 1  $\mu\text{m}$ , hundreds of millions of sequencing reads can potentially be obtained in parallel by rastered imaging of a reasonably sized surface area. (3) Because array features are immobilized to a planar surface, they can be enzymatically manipulated by a single reagent volume. Although microliter-scale reagent volumes are used in practice, these are essentially amortized over the full set of sequencing features on the array, dropping the effective reagent volume per feature to the scale of picoliters or

**Table 6.2** Comparison of next-generation sequencing platforms

Platform	Library/template preparation	Sequencing principle or NSG chemistry	Read length (bp)	Strength	Weakness
454	Fragment, Mate-pair, emulsion PCR	Polymerase (Pyrosequencing)	200–300	Longest read length, fast run time	Asynchronous, homopolymers (e.g., CCC, AAA) enhances error rate, high reagent cost
Solexa	Library consists of a mixture of adapter flanked fragments, uses bridge PCR	Polymerase (Reversible terminator)	75–100	Most widely used platform	Multiplex capacity of the sample is low
AB SoliD	Libraries consist of a mixture of short adapter flanked fragments; uses emulsion PCR	Noncleavable probe, cyclic sequencing by ligation (octamer with two base encoding)	50	Error can be corrected by ligase enzymology, primer reset and “two-base encoding”	Read length is low, run-time is long
Polonator	Mate-pair, uses emulsion PCR	Cyclic sequencing by ligation (nonamer)	26	Relatively cheaper; can be adjusted to other NGS chemistry	Shortest read length
HelixScope	Fragment, Mate-pair, Single molecule	Reversible terminator principle	25–32	No clonal amplification is needed; problem of homopolymer can be tackled.	Low read length, higher error rate

femtoliters. Collectively, these differences translate into dramatically lower costs for DNA sequence production (Shendure and Ji 2008).

The disadvantages of second-generation DNA sequencing includes (1) shorter read-length, i.e., read-lengths are currently much shorter than conventional sequencing, (2) low accuracy, i.e., base-calls generated by the new platforms are less accurate than those of Sanger sequencing. In times to come, these parameters will be taken care of to make it more effective in terms of quantity as well as quality.

The important applications of next-generation sequencing include (1) full-genome resequencing, (2) mapping of structural rearrangements, (3) “RNA-Seq,” analogous to expressed sequence tags (EST) or serial analysis of gene expression (SAGE), (4) large-scale analysis of DNA-methylation, by deep sequencing of bisulfite-treated DNA; and (5) “ChIP-Seq,” or genome-wide mapping of DNA–protein interactions, by deep sequencing of DNA fragments pulled down by chromatin immunoprecipitation. However, many more new applications are expected in the coming days (Shendure and Ji 2008).

### 6.4.3 *Gene Discovery and Deployment*

Discovery and deployment of a useful gene is a major scientific challenge determining the success of developing plant varieties suitable for climate-resilient agriculture. Therefore, several approaches have been put in place for high-throughput gene discovery. The gene silencing technique that employs double-stranded RNA (dsRNA)-mediated interference of functional genes has been effectively and widely used in plants. Similarly, the virus-induced gene silencing (VIGS) technique has also been used to knockout endogenous genes in a transient manner. In VIGS, the vector virus carries a sequence from the plant; the transcript of both the viral and the homologous endogenous gene are degraded by the posttranscriptional gene silencing (PTGS) mechanism (Kumagi et al. 1995; Burton et al. 2000; Ratcliff et al. 2001). However, these techniques are not smooth functioning in plant systems; transformation efficiency is a primary factor of concern. Gene knockout through transposon and T-DNA insertion were also tested in large scale in plant systems for functional annotation of genes (Krysan et al. 1996). Several insertion mutant populations were developed and gene function could be assigned through analysis of mutant phenotype (Winkler et al. 1998). Although it is a powerful technique, the degree of target specificity is a factor to be considered; sometimes it fails to knock the gene of interest. It might not be so effective in species having more than one genome such as soybean, a paleopolyploid. Targeting induced local lesion in genome (TILLING; McCallum et al. 2000) is the other effective and high-throughput reverse genetics approach for identification and study of mutation in plants. It has been successfully utilized in maize and soybean (Till et al. 2004; Slade et al. 2005; Mizoi et al. 2006; Horst et al. 2007). Such approaches would be of



immense help in identifying useful genes for deployment in breeding programs to develop crop varieties suitable for changed climatic conditions.

## 6.5 Modern Plant Breeding Approach

### 6.5.1 *Molecular Plant Breeding*

Plant breeding is as old as agriculture. However, it as a branch of science and field of research received attention and importance only after Mendel's principles of character transmission were rediscovered and realized i.e., after 1900 AD (Allard 1960). With the understanding of how genes control different traits of importance and become transmitted from one generation to another, plant breeders started incorporating gene(s) of choice to the desired genetic backgrounds through different hybridization and selection techniques. However, genes are not independent of the environment (external and internal), and often show their impact through interactions. The understanding of statistics and biometrical genetics helped in unraveling the role of the major and minor genes and the environment, and showed a way to manipulate such traits i.e., the quantitative traits. The locus that governs such quantitative traits is often called the quantitative trait locus (pl. loci) (QTL, in short). Using genetic principles in breeding, breeders succeeded in enhancing the yield of crops as well as crops' capacity to resist disease and insect-pest attacks to a great extent. With deployment of specific alleles of genes, yield of wheat and rice were enhanced to such an extent that it saved millions of poor people from hunger, particularly in Asia and Africa. This success of plant breeding in the 1970s was christened the "green revolution," and the key person behind it, Norman E. Borlaug, was conferred with a Nobel Prize for Peace. Thus, the success of modern plant breeding is unquestionable and well documented (Fehr 1984). It is difficult to believe that such a feat would have been achieved without the clear understanding of the principles of plant breeding and genetics.

With the rise of global population and corresponding increase in eco-environmental problems, the plant breeding techniques demand modernization for enhanced efficiency and environment suitability. Presently, the techniques of plant breeding need to be such that they deliver hugely in the fastest possible time and with the least harm to ecosystems. This led to the introduction and application of molecular biological tools and techniques for modernization of conventional plant breeding, now called "molecular plant breeding" or "molecular breeding," for short. The last three decades have witnessed the development and use of a number of molecular markers, including RFLP, RAPD, AFLP, SSR, and SNP, etc., for various purposes of crop breeding and experimentations. These markers have enabled the breeders to select the desired plant from a large segregating population in an efficient and environment-independent way. Because of its effectiveness, molecular marker-assisted breeding has become equivalent to the conventional

breeding process for many giant agricultural companies including Monsanto, Pioneer, etc (Paul 2009). Markers have made it possible to breed traits from otherwise discarded varieties back into cultivated crops. The conventional backcrossing program has become more potent with the addition of molecular markers, nowadays called the marker-assisted backcrossing (MAB) program. The success of this technique has been recognized by the transfer of many important traits, disease and insect resistance in particular, into crops including rice, maize, soybean, etc. The most prominent success has been the development of submergence-tolerant rice, which was developed by transferring the submergence-tolerance gene (QTL) (*sub-1*) from an Indian landrace to an improved cultivar (e.g., Swarna) through MAB. Thus, molecular breeding has empowered the breeders to tackle even QTL, which the conventional plant breeder found difficult to handle because of its complex genetics and sensitivity to environmental fluctuations. With the availability of a large number of markers, efficient computing power and incorporation of modern concepts like genomic selection (GS), it has become possible to make breeding decisions based on every gene influencing a trait, not just a few. Although molecular breeding has a long way to go to meet the goal of feeding the millions of people on earth, its promise has begun to catch the hype.

### 6.5.1.1 Allele Mining

Allele mining is an approach to identify naturally occurring allelic variants at loci of agronomic importance, i.e., those genes that affect crop characteristics and performance. In this sense, it includes analysis of noncoding and regulatory regions (“promoter mining”) of the candidate genes in addition to analyzing sequence variations in the coding regions of the agronomically important genes (Rangan et al. 1999; Latha et al. 2004). The alleles of genes that are candidates for the target trait can be identified using a variety of approaches including mutant screens (Johal and Briggs 1992; Whitham et al. 1994; Bishop et al. 1996), QTL analysis (Backes et al. 1995; Xiao et al. 1996; Bernacchi et al. 1998), association mapping (Crossa et al. 2007; González-Martínez et al. 2007), and genome-wide surveys for the signature of artificial selection (Vigouroux et al. 2002; Casa et al. 2005; Yamasaki et al. 2005; Chapman et al. 2008). Such novel and important alleles recovered at loci of agronomic importance can be integrated into crop breeding programs using conventional or molecular approaches for combating biotic and abiotic challenges, to enhance yield, improve storage and nutritional qualities. However, success of allele mining operations is dependent on the availability of diverse germplasm collections (Kumar et al. 2010), as wild relatives of the crops are the repository of the useful new alleles not already present in the crop gene pool (Tanksley and McCouch 1997; Gur and Zamir 2004; Johal et al. 2008; Prada 2009).

Two approaches that are generally utilized for identification of sequence polymorphisms for a given gene in the naturally occurring populations includes (1) modified TILLING and (2) sequencing-based allele mining (Kumar et al. 2010). TILLING (McCallum et al. 2000) is a reverse genetics approach that can identify

point mutation in a target gene by heteroduplex analysis (Till et al. 2004). TILLING or its modified version Eco-TILLING is based on the enzymatic cleavage of heteroduplex DNA formed due to mismatch in sequence between reference and test genotype with a single-strand specific nuclease (e.g., Cel-1, S1, mungbean nuclease, etc.) under specific conditions followed by detection through Li-Cor genotyper or capillary electrophoresis (CE) separation. At the site of point mutation, there will be cleavage by the nuclease to produce two cleaved products whose sizes will be equal to the size of the full-length product. The presence, type, and location of point mutation or SNP will be confirmed by sequencing the amplicon from the test genotype that carries the mutation. Conversely, the sequence-based TILLING approach involves amplification of alleles in diverse genotypes through PCR followed by identification of nucleotide variation by DNA sequencing. Both approaches are not cost-effective, as claimed. However, with the advent of newer sequencing technologies, the second- and third-generation sequencing technologies in particular, that are capable of producing at low cost, with more read length and high throughput, sequence-based allele mining is expected to generate data at an economically affordable level.

The most important application of allele mining is the discovery of superior alleles from unutilized natural plant genetic resources. A number of such alleles has already been identified and used in crop improvement. It can also be used to reason the molecular basis of novel trait variations and pinpoint the nucleotide sequence changes associated with superior alleles. It may pave the way for molecular discrimination among related species, development of allele-specific molecular markers, facilitating introgression of novel alleles through MAS or deployment through genetic engineering (Kumar et al. 2010).

### 6.5.1.2 Quantitative Trait Locus Mapping

The QTL-mapping approach generally begins with the crossing of two parental inbred lines for a number of generations to form preferably a population of recombinant homozygous lines. There are several methods for QTL mapping ranging from the simplest method of single-marker analysis (Sax 1923) to more sophisticated methods such as interval mapping (Lander and Botstein 1989; Haley and Knott 1992), joint mapping (Kearsey and Hyne 1994), multiple regression (Wright and Mowers 1994; Whittaker et al. 1996), and composite interval mapping (Zeng 1994). Analytical complexities are taken care of with various software packages including MAPMAKER/QTL (Lincoln et al. 1993), JoinMap (Stam 1993), QTL Cartographer (Basten et al. 1994), PLABQTL (Utz and Melchinger 1996), QGene (Nelson 1997), and TASSEL (Buckler 2007). This approach has proved to be quite useful for plant breeding and has been successful in identifying loci of large effect and dissecting the genetic basis of fairly simple traits. A detailed discussion of QTL analysis has been given in Chap. 4.

The QTL approach will be important in terms of climate-resilient agricultural activities. Most of the traits particularly important in the event of climate change are

quantitative in nature. Moreover, QTLs for tolerance to traits including drought, heat, frost, flood, etc., have been reported in different genetic backgrounds. Now, accumulating useful QTLs is the major goal for crop improvement in the changing climatic situation. In this regard, various approaches as described in this chapter will be the foci of the future for climate-resilient plant breeding. QTL mapping for a drought-tolerance trait has been done in different crops including maize, wheat, barley, cotton, sorghum, and rice (Quarrie et al. 1994; Teulat et al. 1997; Sari-Gorla et al. 1999; Saranga et al. 2001; Sanchez et al. 2002; Bernier et al. 2008). In pearl millet, a major QTL associated with grain yield in drought stress environments has been identified on linkage group 2 (LG 2), which accounts for up to 32 % of the phenotypic variation of grain yield (Yadav et al. 2011).

The primary disadvantages of the QTL mapping approach is the time involved in creating populations, limited information and inferences that can be made from alleles in two parental lines, the small number of recombination events captured in most mapping populations and a necessary focus on traits that can be readily and accurately phenotyped (Morrell et al. 2012). To overcome such limitations and to map QTL with higher precision, approaches like marker-assisted recurrent selection (MARS) and genome-wide selection (GWS) have been used for such types of study (Ravi et al. 2011).

### 6.5.1.3 Association Mapping

Association or linkage disequilibrium (LD) mapping approaches assess the correlation between phenotype and genotype in populations of unrelated individual lines. The association-mapping panels (Risch and Merikangas 1996) sample more genetic diversity, can take advantage of many more generations of recombination, and avoid the generation of time-consuming crosses that are necessary for QTL mapping (Myles et al. 2009). Unlike QTLs identified through biparental mapping strategies that can span tens of megabases, the recombination event that is captured in most association panels enables a much greater genetic resolution. With a large panel and with sufficiently dense genome-wide marker coverage, association mapping can potentially map causative loci to individual nucleotide changes. However, much higher resolution can be achieved through genome-wide association studies (GWAS). Huang et al. (2010) used low-coverage resequencing of the genomes of a panel of more than 500 rice landraces and found 80 loci associated with 14 agronomic traits. Several of these associations were previously characterized showing authenticity of the results. Precision of similar kind was also reported in *A. thaliana* (Atwell et al. 2010). These results suggest limits to the precision available in association mapping studies, particularly in inbreeding organisms (Atwell et al. 2010; Hamblin et al. 2011).

### 6.5.1.4 Marker-Assisted Recurrent Selection

The plant breeder usually uses a conventional recurrent selection process in order to increase the frequencies of desirable QTL alleles at multiple loci. Now, two related approaches have been proposed for similar purposes (1)  $F_2$  enrichment followed by inbreeding (Howes et al. 1998; Wang et al. 2007); and (2) MARS (Edward and Johnson 1994; Hospital et al. 1997; Johnson 2004; Bernardo and Charcosset 2006). In both these approaches, the starting material usually is an  $F_2$  population. The primary goal is to develop an improved recombinant inbred (RI) in the case of self-pollinated crops and with superior testcross performance for cross-pollinated crops. In  $F_2$  enrichment, the  $F_2$  plants with undesirable alleles are culled in order to increase the frequency of desirable alleles that in another way increases the probability of recovering inbred with desirable alleles at all the loci. Usually, this selection cycle is run only once, because a second or third round of enrichment adds little advantage (Wang et al. 2007). However, in MARS multiple cycles of selection are performed based on markers (Edward and Johnson 1994; Johnson 2004; Eathington et al. 2007). The step in MARS involves (1) identifying  $F_2$  plants or  $F_2$ -derived progenies that have the desirable allele at most of the target QTLs, (2) recombining selfed progenies from these selected individuals, and (3) repeating the procedure for 2–3 cycles. In the MARS approach, a selection index is created based on weight given to markers according to the relative magnitude of their estimated effects on the trait (Lande and Thompson 1990; Edward and Johnson 1994). The selection index is given by  $M_j = \sum b_i X_{ij}$ , where  $M_j$  is the marker score of the  $j$ th individual,  $b_i$  is the weight given to the  $i$ th marker locus and  $X_{ij}$  is given score 1 if the  $j$ th individual is homozygous for the marker allele with favorable effect, and  $-1$  if the individual is homozygous for the marker allele with un/less favorable effect. The value of the  $b_i$  weight can be obtained from multiple regressions of trait values on  $X_{ij}$  (Lande and Thompson 1990; Hospital et al. 1997). MARS is now being used in several crops. Success has been reported in sweet corn where MARS increased the frequency of the favorable alleles from 0.50 to  $\geq 0.80$  at 18 out of 31 markers used in selection (Edward and Johnson 1994), and five marker loci became fixed for the favorable alleles. Similarly, in another  $F_2$  population, the frequency of the favorable marker allele increased to  $\geq 0.80$  at 11 out of 35 markers used in selection. This shows the power of the approach. It has now been tried in other crops including wheat and success is underway. A possible disadvantage of MARS is the necessity of extra numbers of generations for cyclic selection based on markers (Bernardo 2008).

### 6.5.1.5 Genome-Wide Selection or Genomic Selection

GWS refers to marker-based selection without significance testing and without identifying a subset of markers associated with the trait (Meuwissen et al. 2001). Here, the effects of all the markers on the quantitative trait (i.e., breeding value) are

fitted as random effects in a linear model. Trait values are then predicted as the sum of an individual's breeding values across all the markers used, and selection is subsequently based on these genome-wide predictions. Therefore, the first step of genomic selection (GS) is the estimation of marker effects and the designing of genomic prediction models from a training population. The second step consists of two or three cycles of early-generation genotypic selection and intercrossing, allowing accumulation of favorable alleles. GWS leads to high correlations between predicted and true breeding value for a quantitative trait (Meuwissen et al. 2001). It has been shown that GS can increase efficiency of breeding for yield potential in elite x elite crosses, but it cannot introduce new genetic variability into the most adapted backgrounds. Trait-based breeding will combine and introgress yield potential traits coming from various genetic resources into the most adapted recipient parent (Reynolds et al. 2011).

#### 6.5.1.6 Next-Generation Mapping of Complex Traits

Mapping complex traits is always difficult and challenging. Various approaches have already been proposed and used with limited success. In recent times, next-generation sequencing (NSG) approaches have also been used for this purpose. Basically, NSG mapping is comprised of three major steps (1) development of the mapping population, (2) extraction of DNA and preparation of libraries, and (3) assembling short reads and genotyping. The NSG approaches have the potential to map QTL more effectively. However, such approaches may also lack sufficient power. Whole-genome resequencing (WGR) (Huang et al. 2009), restriction site-associated DNA (RAD, Baird et al. 2008) and phenotype-based selection and introgression followed by whole-genome resequencing (PSIseq, Earley and Jones 2011) are NSG-based mapping approaches. WGR and RAD use a bulk segregant population and require large mapping populations to detect multiple loci with weak effect (Ehrenreich et al. 2010). In PSIseq, populations with a divergent complex trait are hybridized and then selected for a specific phenotype across multiple generations of backcrosses. The trait of interest is selected for each generation, and offspring are mated to the other parental line expressing the unselected phenotype (introgression and backcrossing). Over multiple generations of selection and backcrossing this hybrid population becomes homozygous for the majority of the unselected parent's genome while loci from the selected parent, which contribute to the selected trait, remain. Using high-throughput sequencing, the breakpoints of introgression are mapped, which eventually map the regions harboring genes influencing the trait. The power and effectiveness of PSIseq have been shown in the *Drosophila* system. However, it can be used in other systems as well. It does not require existing high-quality reference genomes. Unlike other NSG mapping approaches, it can find multiple loci with small mapping populations (Earley and Jones 2011). With the wider applications of NSGs, these mapping approaches will find a popular place in plant breeding programs of the future.

## 6.5.2 Next-Generation Mapping Populations

A multiparent populations concept, called next-generation populations (NGPs), has evolved and been used recently with the aim to overcome the shortcomings of biparental QTL mapping and association mapping populations. In principle, it combines the controlled crosses of QTL mapping with multiple parents and multiple generations of intermating. The NGPs are often larger than traditional QTL populations, and many lines are crossed in parallel; this increases the rate of effective recombination per generation and maximizes “genetic map expansion,” thereby improving genetic resolution compared to traditional biparental mapping (Rockman and Kruglyak 2008). Like association mapping panels, NGPs will more effectively sample rare alleles than typical biparental populations. It also has the power to overcome some of the difficulties of association mapping, including population structure and the unknown frequency of causative mutations. Further, it allows better estimation of allelic effects than is possible under standard association mapping approaches purely because of equal contribution of all the parents involved (Macdonald and Long 2007).

### 6.5.2.1 Nested Association Mapping Population

The nested association mapping (NAM) population is a form of next-generation population. Here, diverse strains are crossed to a reference parent (Yu et al. 2008). The resultant  $F_1$ s are self-fertilized for several generations in order to develop a series of RILs (Fig. 6.2). The members of the RIL families are either sequenced or genotyped and compared with the reference lines. In maize, 25 diverse corn lines were used as the parental lines and each parental line was crossed to the B73 maize inbred to produce an  $F_1$  population. The  $F_1$  plants were then self-fertilized for six generations in order to create a total of 200 homozygous RILs per family, for a total of 5,000 RILs within the NAM population.

NAM combines the advantages and eliminates the disadvantages of two conventional methods for identifying QTLs: linkage analysis and association mapping. In principle, the linkage analysis depends upon the recent genetic recombination while the association mapping takes advantage of historic recombination. Linkage analysis suffers from low-resolution mapping while association mapping needs extensive knowledge of SNPs. The NAM takes advantage of both historic and recent recombination events and, eliminates the disadvantages associated with both of the traditional approaches. It has successfully been utilized in the study of numerous traits in maize (Buckler et al. 2009; Brown et al. 2011; Kump et al. 2010; Tian et al. 2011). The NAM population has tremendous potential to be used in genome-wide association studies and for study of agronomic traits useful for changing climatic conditions.

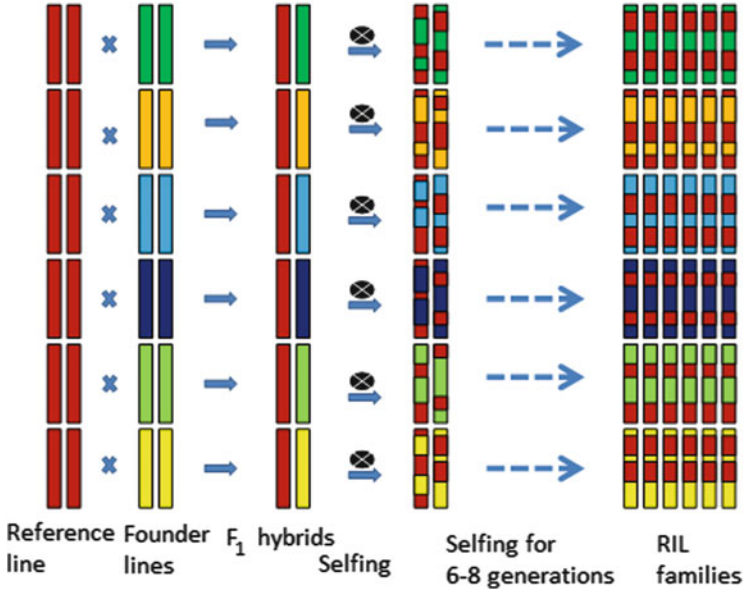


Fig. 6.2 Schematic diagram of NAM population generation

### 6.5.2.2 Multiparent Advanced Generation Intercross Population

The multiparent advanced generation intercross (MAGIC) is another form of next-generation population, which involves intercrossing of multiple parents, forming a single large population (Cavanagh et al. 2008; Kover et al. 2009) (Fig. 6.3). It is also called recombinant inbred advanced intercross line (RIAIL) populations (Rockman and Kruglyak 2008). Multiparent crosses have a long history in plant breeding (Harlan and Martini 1929; Suneson 1956) and have been a source of considerable insight into evolutionary processes in crops (Allard et al. 1972; Clegg et al. 1972). However, adoption of such mapping populations will vary with the crop species depending upon the ease or complexities with which crosses can be made. For example, designs that involve repeated outcrossing (e.g., MAGIC) are difficult to implement in self-pollinated crops such as soybean (Cavanagh et al. 2008).



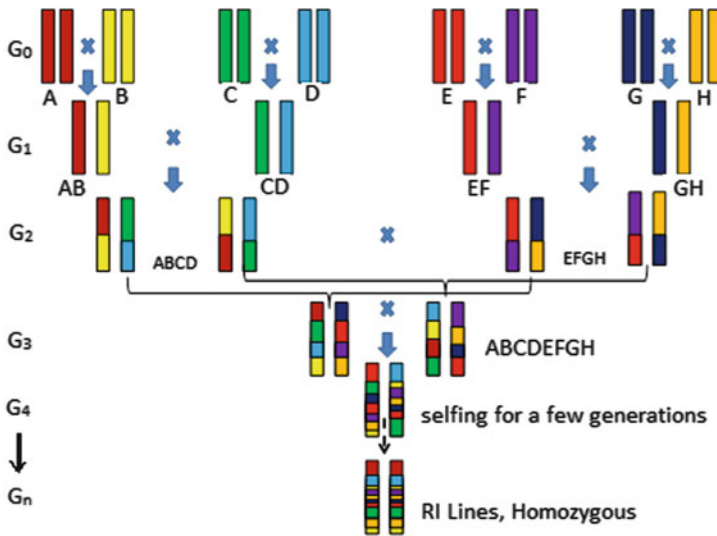


Fig. 6.3 Schematic diagram of MAGIC population generation

## 6.6 Advanced Approaches for Utilization of Plant Genetic Resources

### 6.6.1 Introgression of Desirable QTL Through Wide Hybridization

Wild or weedy relatives of crop species are an important reservoir of agronomically superior traits. Such species possess the genes/alleles that helped it survive in changed climatic conditions over the years. Such wild relatives of crops have contributed several important traits for improvement of current cultivars of rice, wheat, maize, barley, and other crops (Hajjar and Hodgkin 2007). In the recent past, the genetic and cytogenetic approaches have contributed significantly towards such transfer of genes from alien species into various crops including rice (Ashikari et al. 2005) and potato (Maxted et al. 2007). An understanding of its genetic and genomic basis would help in planning breeding programs for transfer of the desirable genes from such species to otherwise elite crop cultivars in a more effective and efficient way. The molecular approaches have the technical ability to overcome problems such as linkage drag, shattering, etc., associated with transfer of desirable alleles from wild species. Several molecular marker-supported approaches have been utilized for efficient transfer of desirable traits from wild species into tomato, barley, rice, and so on. Approaches like introgression line (IL, Eshed and Zamir 1995), single segment substitution lines or 3-S lines (Talukdar and Zhang 2007), chromosome segment substitution lines (CSSLs, Wang et al. 2004), and stepped aligned inbred recombinant strains (STAIRS; Koumproglou et al. 2002) have

shown promise in transfer of traits mostly governed by major genes. However, these can also be used for other traits for enrichment of genetic variations and development of hardy crops that can survive and sustain themselves against the vagaries of nature.

### **6.6.2 Advanced Backcross QTL Analysis**

It has been proved time and again that despite the inferior phenotype, the exotic germplasm contain QTLs that can increase the yield and quality of elite breeding lines. However, these QTLs are often linked with undesirable traits making their application difficult. Efforts to overcome such hurdles have motivated the development of a new molecular breeding strategy, referred to as the advanced backcross (AB) QTL method (Tanksley and Nelson 1996), which integrates QTL analysis with variety development, by simultaneously identifying and transferring favorable QTL alleles from unadapted to cultivated germplasm. The AB-QTL strategy has so far been tested in tomato, rice, barley, maize, pepper, etc. In this approach, QTL and marker analyses are performed in advanced generations, like BC<sub>2</sub> or BC<sub>3</sub>. During the development of these populations, a negative phenotypic and/or genotypic selection pressure is exerted against the unfavorable genes particularly originating from the donor parent. This helps in the reduction of the percentage of the donor-parent genome simultaneously reducing the alleles that could otherwise interfere with yield and other field performance traits (Tanksley and Nelson 1996). In the case of cross-pollinated crops, field testing can be conducted on BC<sub>2</sub>S<sub>1</sub> or BC<sub>3</sub>S<sub>1</sub>. Conversely, for crops where commercial hybrids are more commonly used, the BC<sub>2</sub> or BC<sub>3</sub> plants are crossed with a tester variety to generate BC<sub>2</sub>F<sub>1</sub> or BC<sub>3</sub>F<sub>1</sub> families. Once favorable QTL alleles are identified, only a few more marker-assisted generations are required to develop near-isogenic lines (NILs) that can be field tested and used for variety development. Therefore, a cycle of AB-QTL analysis (i.e., QTL discovery, NIL development and testing) represents a direct test of the underlying assumption of QTL breeding: that beneficial alleles identified in segregating populations (i.e., BC<sub>2</sub> or BC<sub>3</sub> in the case of AB-QTL) will continue to exert their positive effects when transferred in the genetic background of elite lines (Grandillo and Tanksley 2003). This process not only results in improved elite varieties, but it also provides a strategy for selectively expand the genetic base of crop species, especially those with a narrow germplasm base. In the effort to breed crops for changing climatic conditions, this approach might prove to be extremely useful as it offers a way to transfer the desirable QTLs from the wild and weedy species to the otherwise elite genotypes, which might lack the gene(s) for adaptability to the harsh environments that are predicted to occur in the years to come. Further, use of high-density molecular marker maps would ensure transfer of the desirable QTLs in a more precise manner making the fear of linkage drag remote.

### 6.6.3 Association Mapping

Association mapping, also known as linkage disequilibrium (LD) mapping is a powerful tool to resolve complex trait variation down to the sequence level by exploiting historical and evolutionary recombination events at the population level (Risch and Merikangas 1996; Nordborg and Tavaré 2002). As compared to traditional linkage mapping analysis, the advantages of association mapping are (1) increased mapping resolution, (2) reduced research time, and (3) greater allele number (Yu and Buckler 2006). Basically, there are two approaches for association mapping (1): a candidate-gene based approach, which relates polymorphisms in selected candidate genes that have purported roles in controlling phenotypic variation for specific traits; and (2) a genome-wide or genome scan approach, which surveys genetic variation in the whole genome to find signals of association for various complex traits (Risch and Merikangas 1996). For the candidate-gene approach, information regarding the location and function of the genes involved in either biochemical or regulatory pathways are required, which are currently available for most of the model plants and a number of other crop species, as well. At the same time, the availability of hundreds of thousands of SNPs at affordable cost through resequencing has prompted researchers to move toward genome-wide association analyses of complex traits. The *Arabidopsis* HapMap provided a powerful catalog of genetic diversity with more than one million SNPs at an average rate of one SNP every 166 bp (Clark et al. 2007). The unique advantage of association mapping is that it harnesses the genetic diversity of natural populations to potentially resolve complex trait variation to single genes or individual nucleotides, and hence its results have wider applicability.

Linkage disequilibrium or gametic phase disequilibrium is the measurement of the degree of nonrandom association between alleles at different loci. The difference between observed haplotype frequency and expected based on allele frequencies is defined as  $D$ . It is given as,  $D = p_{AB} - p_A p_B$ , where  $p_{AB}$  is the frequency of gamete AB;  $p_A$  and  $p_B$  are the frequency of allele A and B, respectively. In the absence of other forces, recombination through random mating breaks down the LD with  $D_t = D_0(1 - r)^t$ , where  $D_t$  is the remaining LD between two loci after  $t$  generations of random mating from the original  $D_0$ . Several statistics have been proposed for LD, and these measurements largely differ in how they are affected by marginal allele frequencies and small sample sizes (Hedrick 1987). The estimates of both  $D'$  (Lewontin 1964) and  $r^2$  (Hill and Robertson 1968) have been widely used to quantify LD. In terms of identifying SNPs or haplotypes significantly associated with phenotypic trait variation,  $r^2$  is the most relevant LD measurement. Typically,  $r^2$  values of 0.1 or 0.2 are often used to describe the LD decay. Though LD is affected by many factors (Ardlie et al. 2002), LD due to linkage is the net result of all the recombination events that occurred in a population since the origin of an allele by mutation, providing a greater opportunity for recombination to take place between any two closely linked loci than that in linkage analysis (Holte et al. 1997; Karayiorgou et al. 1999). Among other factors,

the reproduction mode of a species partly determines the level of LD in a diverse population (Flint-Garcia et al. 2003). Generally, LD extends to a much longer distance in self-pollinated crops, such as soybean, wheat, than in cross-pollinated species, such as maize. The shorter the LD decay, the greater the mapping resolution, and vice versa.

Association or LD mapping often encounters problems of spurious association generated by population structure and familial relatedness (Yu and Buckler 2006), which needs to be taken care of through statistical treatments. Several statistical approaches viz., structured association, genomic control, mixed model, principal component, etc., can be adopted for explicit or ad hoc adjustment (Yu et al. 2006).

The size of the sample taken for association mapping remains relatively small. In many recent association mapping studies, only about 100 lines were investigated. However, unless the functional locus has a very large effect and tested markers are in high LD with this locus, it will be difficult to identify marker–trait associations with a small population, regardless of whether the candidate-gene or genome-scan approach is used. Simulations with empirical maize data showed that a large sample size is required to obtain high power to detect genetic effects of moderate size (Zhu et al. 2007). Similarly, the number of markers to be used is also an important consideration for candidate-gene association mapping studies. Generally, the number of required markers is much higher for biallelic SNPs than for multiallelic SSRs. As a starting point, the number of SSR markers should be about four times the chromosome number of that species, i.e., at least two markers per chromosome arm. However, chromosome length, diversity of the species and the particular sample, cost and availability of different marker systems will be the ultimate determining factors.

A variety of software packages are available for data analysis in association mapping (Table 6.3). However, TASSEL and STRUCTURE are the most commonly used software for association mapping in plants. TASSEL is also used for calculation and graphical display of linkage disequilibrium statistics and browsing and importation of genotypic and phenotypic data. STRUCTURE software typically is used to estimate the Q matrix (Pritchard et al. 2000). However, to resolve specific issues, specialized software has to be used.

Association mapping or LD mapping was initially developed in human beings. However, it has an equal or even better foothold in animal and plant systems. It has successfully been used in almost all the major crops such as maize, soybean, barley, wheat, tomato, sorghum, and potato, as well as tree species such as aspen and loblolly pine, to name a few. With the availability of the genome sequences of the important crop species and sophisticated analytical tools, the association mapping approach is expected to provide more useful information to breeders allowing them to develop designer crops with the capability to sustain themselves during global climate change. This approach has been advanced further incorporating the benefits of both linkage mapping and association mapping, in an approach called NAM. In this approach, a large number of lines (founder lines) are crossed to a single genotype (reference line). The  $F_1$  thus produced are selfed for 6–8 generations to generate a series of RIL families (Fig. 6.2). This multiparent mapping population is

**Table 6.3** List of a few software packages commonly used for components of association analysis

Software	Source	Remarks
ASREML	<a href="http://www.vsnr.co.uk/products/asreml">http://www.vsnr.co.uk/products/asreml</a>	Basically designed for animal breeding data analysis. However, it can be used for plant breeding data. It uses a mixed model analysis system
EINGENSTRAT PCA	<a href="http://genepath.med.harvard.edu/~reich/Software.htm">http://genepath.med.harvard.edu/~reich/Software.htm</a>	It is used for association analysis, PCA
MTDFREML	<a href="http://aipl.arsusda.gov/curtvt/mtdfrem1.html">http://aipl.arsusda.gov/curtvt/mtdfrem1.html</a>	It uses mixed model analysis of data primarily of animal breeding, can also be used for plant breeding data
R	<a href="http://www.r-project.org/">http://www.r-project.org/</a>	A generic software convenient for simulation work
SPAGeDi	<a href="http://ebe.ulb.ac.be/ebe/SPAGeDi.html">http://ebe.ulb.ac.be/ebe/SPAGeDi.html</a>	It is fit for genetic and relative kinship analysis
SAS	<a href="http://www.sas.com">http://www.sas.com</a>	It is also a generic software package used widely for data analysis
STRUCTURE	<a href="http://pritch.bsd.uchicago.edu/structure.html">http://pritch.bsd.uchicago.edu/structure.html</a>	It is widely used for analysis of population structure
TASSEL	<a href="http://www.maizegenetics.net">http://www.maizegenetics.net</a>	It is used for association analysis, analysis of LD; has a component of linear and mixed model for data analysis

gaining popularity among breeders for its applicability to creating the wider variability necessary in improving crop species for wider adaptation.

#### 6.6.4 Multiparent Advanced Generation Intercross

Most of the traits of agronomic importance are polygenic, i.e., controlled by polygenes. Such traits are difficult to analyze for their effects and inheritances. Since these traits are governed by polygenes each with small, similar and additive effect, they hence show a continuous variation in the segregating populations. Most importantly, the effects of such genes are not pronounced enough and hence are highly influenced by the environment making estimation of genetic effects more complex. Therefore, complex statistical and biometrical approaches are applied in the analysis of such traits (commonly called quantitative traits). Genetic variations expressed by such traits are an important source for genetic analysis. For genetic studies and, more particularly for molecular mapping and analysis of QTLs, a series of synthetic population, viz., backcross, RILs, and doubled-haploid (DH) lines are used. Usually, such populations (except DH) are developed by crossing two diverse genotypes. The  $F_1$  generation is either selfed (to produce  $F_2$ ) or backcrossed with one of the parents (usually the recurrent parent) to produce backcross populations. The resultant generations starting from  $F_2$ s may be advanced through the single

seed descendent (SSD) approach to produce RIL. These populations, called mapping populations have their own niche and limitations. The most serious limitation is the limited variability i.e., the variability is limited to present in the two parents only. Further, the resolutions with which the QTLs are mapped are statistically poor. QTLs identified by these mapping populations have confidence intervals of 5–20 cM (Wilson et al. 2001; Louder et al. 2002; Ungerer et al. 2002), which corresponds on average to 1.2–4.8 Mb (million base pairs) covering hundreds of candidate genes. To overcome such limitations, a series of populations have been developed by various workers over a period of time. The unique feature of these new generation mapping populations is that they involve more than two genotypes (parents) in their population, and are amenable to replications over time and place that improves the power and resolution of QTL mapping significantly. One such mapping population is MAGIC (Kover et al. 2009) (Fig. 6.3). In this population, a number of accessions (founder lines) are intermated in a complete diallel fashion i.e., each accession is crossed with all other accessions as a maternal and paternal parent. The  $F_1$  thus produced are then intermated through four generations of random mating. From the  $F_4$  families, 2–3 MAGIC lines (MLs) are derived by selfing for four to six generations. Precautions are to be taken to avoid assortative mating through staggered planting or planting the same families multiple times (see Scarcelli et al. (2007) for the detail procedure). The MLs thus derived are subjected to phenotyping in replicated trials and genotyping with molecular markers for detection of SNPs. Through appropriate statistical treatment and uses of analysis software, QTLs are detected with higher resolution. This approach does not necessitate repeated genotyping and allows study of trait correlations, genotype-by-environment interactions, and the genetic basis of phenotypic plasticity. As it is developed through several intercrossings, and can be replicated unlimitedly, associations are located with greater accuracy and, QTLs are mapped with higher resolution. This has successfully been utilized to map QTLs for bolting time, days to germinate, and glabrous and erecta traits in *A. thaliana* (Kover et al. 2009). MAGIC populations are being developed in the UK and CSIRO, Australia to incorporate a large proportion of the genetic diversity within elite wheat varieties from around the world. It also enables the discovery of the best combinations of genes for important traits. CSIRO has developed two MAGIC populations. The first is a four parent population and includes lines adapted to all breeding regions in Australia, with genetic diversity covering approximately 80 % of Australian material. The second population is an eight parent population and includes cultivars from six countries (Australia, Canada, China, Israel, United States, and Mexico) (CIMMYT) (CSIRO 2011). Such MAGIC populations would enable genetic dissection of the QTLs and their deployment for the genetic improvement of crops.

## 6.7 Specific Issues for Future Breeding Activities

Breeding crop varieties especially suitable for climate-resilient agriculture should invariably consider a number of climatic factors in addition to the efficiency of the crops in utilizing energy, water, nutrients as well as radiation.

**Radiation Use Efficiency.** Basic research shows that crop productivity and adaptability can be increased many fold through manipulation of photosynthetic mechanisms. The  $C_4$  plants (example, maize) are more efficient than the  $C_3$  plants (example, rice) owing to their high carbon fixation efficiency. The increased concentration of atmospheric  $CO_2$  owing to global air pollution has become beneficial to the  $C_3$  plants. It is estimated that 0.3 % of the observed 1 % rise in global wheat production can be attributed to this increase (Fisher and Edmeades 2010). Reynolds et al. (2011) mentioned that crop productivity increase must achieve a number of broad objectives simultaneously (1) increase of crop biomass through modification of radiation use efficiency (RUE), (2) improvement of targeted adaptation of reproductive processes to major crop agro-ecosystems thereby permitting increases in RUE to be translated to grain weight; and (3) enhancement of the structural characteristics of the crop plants to ensure grain yield potential without sacrificing quality due to lodging. However, a multidisciplinary approach is needed to achieve the targets of crop genetic improvement in terms of its suitability to changing climatic conditions.

**Nitrogen Use Efficiency.** Usually, high-yielding varieties and hybrid crops are nutrient demanding; they need more chemical fertilizer, timely supply of water in addition to chemical protection from time to time. Over or indiscriminate application of chemical fertilizer, nitrogenous fertilizer in particular, has become a matter of concern in the light of water and environmental pollution. It has emerged as a key target in the changing climatic conditions (Tester and Langridge 2010). Therefore, for current breeding programs, it is important to develop crop varieties or hybrids that (1) need lesser amounts of nitrogenous fertilizer, and (2) are highly efficient in their use of the nitrogenous fertilizer. Enhancing or incorporating a biological nitrogen fixation (BNF) capability to the crop varieties could be an important option. Efforts in this direction are ongoing with encouraging progress reported.

**Breeding for Zero Tillage Systems.** The zero tillage approach is gaining popularity among crop producers as a conservation measure for the earth. It causes less damage to the soil structure and thus saves energy needed for field preparation to a greater extent. Its adoption and large-scale use is going to change the spectrum of diseases and insect pests attacking the crops. For such situations, newer breeding strategies have to be designed as the conventional approaches would hardly work. Similarly, breeding approaches have to be changed to meet the need of multiple cropping systems and cropping sequences, for which interactions between agronomists and the breeders is very important.

**GM Crops.** The production and evaluation of genetically modified (GM) crops, which permit generation of variations that may or may not be available in the natural population, will remain an active area of research in the near future. The power and success of this technology has been proved by the significant control of insect pests through deployment of a gene for proteinaceous toxins from the bacterial genome. Likewise, development of  $\beta$ -carotene-rich rice has brought hope for the technology as well as to humanity (Mayer et al. 2008). GM crops with tolerance to abiotic and biotic stresses have been developed in different parts of the world and are in different phases of testing. GM genotypes with tolerance to heat, drought, frost, etc., will be the focus of the future. However, the future focus should be the discovery and characterization not only of genes but also of promoters that provide accurate and stable spatial and temporal control of the expression of the genes (Moller et al. 2009). Similarly, development of marker-free transgenic crops should be on the agenda to avoid political obstructions. The GM technology has the promise to support plant breeding activities for climate-resilient agriculture and will inevitably be deployed for most major crops in the future (Tester and Langridge 2010).

**Targeted Genome Editing Technology.** Transfer of genes/QTLs from unadapted genotypes to an elite genotypic background is often expensive, time consuming, and also suffers from the problem of linkage drag. However, recently developed technologies viz., zinc finger nucleases (ZFN) (Weinthal et al. 2010) and transcription activator-like effector (TALE) nucleases (Bogdanove and Voytas 2011) have the potential to tackle such problems. These technologies involve the use of sequence-specific designer nucleases that cleave targeted loci, enabling creation of small insertions and deletions (indels), insertion of novel DNA, or even replacement of individual alleles. The usefulness of these techniques has been established even in crops (Shukla et al. 2009; Morbitzer et al. 2010). Perhaps, it would not be impossible to target the replacement of deleterious mutations in elite breeding lines (Morrell et al. 2012). Such technologies should open up a path to develop crops that are suitable for the changing climatic situations.

## 6.8 Conclusion

To be precise, there is no single solution to the problem of climate change; nor is there any single plant breeding approach to develop crop varieties capable of countering all climatic hazards that are envisaged as a consequence of changing climatic variables. However, the process of breeding climate-resilient crops would include a few obvious considerations (1) generation of variability: the variants with desirable traits can either be searched for in the available germplasm or created through hybridization, mutation, or transgenic approaches; (2) testing, analysis, and identification: the variability generated or collected needs to be tested for its fitness to the changed climatic variables. For this matter, artificial screening facilities or



phenomic facilities would be useful. The improved plant breeding approaches including molecular tools and techniques, and the genomic approaches would be of great use to identify the target plant or progenies; (3) human resource development: imparting training to research personnel on modern tools and techniques needs to be an important component of such efforts; (4) utilization: finally the tested genotypes should be multiplied and distributed to the end users for wide application. For large-scale adoption, demonstrations can help. In India, such an effort has already been initiated (National Initiative for Climate Resilient Agriculture, NICRA) as a preparatory step towards containing the hazards of climate change in agriculture. Such governmental initiatives and policy support mechanisms are necessary to make the technologies available and accessible to even poor users across the world. In this regard a global initiative is required.

It can be said that a coordinated approach of conventional and modern plant breeding processes coupled with crop management and policy planning can find ways to develop crop varieties and agricultural techniques that can increase crop production and sustain it even under severe environmental conditions. While doing so, the necessity should be to identify and exploit effective alleles at multiple loci through appropriate phenotypic, genotypic, and molecular approaches of selection (Christopher et al. 2007; Crossa et al. 2007). Secondly, to accumulate and utilize greater numbers of favorable alleles, approaches like MARS and genomic selection (GS) should be given due consideration. GS has shown savings of 13 years in the release of improved germplasm in oil palm as compared to conventional approaches (Wong and Bernardo 2008). Such approaches with power, precision, and efficiency should be the foci for crop breeding in climate-resilient agriculture. Thus, it is clear that more is required than what can be provided by traditional breeding approaches alone. Policies should be in place worldwide to make the technologies economically affordable by all sections of users globally. Thus, new genomic and breeding technologies must be developed and employed to accelerate the breeding processes so as to meet the goal of feeding billions of people on earth.

## References

- Allard RW (1960) Principles of plant breeding. Wiley, New York, NY
- Allard RW, Kahler AL, Weir BS (1972) The effect of selection on esterase allozymes in a barley population. *Genetics* 72:489–503
- Ardlie K, Kruglyak L, Seielstad M (2002) Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* 3:299–309
- Ashikari M, Sakakibara H, Lin S, Yamamoto T, Takashi T et al (2005) Cytokinin oxidase regulates rice grain production. *Science* 309:741–745
- Atwell S, Yu S, Huang, Bjarni J, Vilhjálmsson, Willems G et al (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465:631–637
- Backes G, Graner A, Foroughi-Wehr B, Fischbeck G, Wenzel G, Jahoor A (1995) Localization of quantitative trait loci (QTL) for agronomic important characters by the use of a RFLP map in barley (*Hordeum vulgare* L.). *Theor Appl Genet* 90:294–302

- Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL et al (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3:e3376
- Baldwin D, Crane V, Rice D (1999) A comparison of gel-based, nylon filter and microarray techniques to detect differential RNA expression in plants. *Curr Opin Plant Biol* 2:96–103
- Basten CJ, Weir BS, Zeng ZB (1994) QTL Cartographer. Department of Statistics, North Carolina State University, Raleigh, NC
- Baulcombe DC (1999) Fast forward genetics based on virus-induced gene silencing. *Curr Opin Plant Biol* 2:109–113
- Bernacchi D, Beck-Bunn T, Eshed Y, Lopez J, Petiard V, Uhlig J, Zamir D, Tanksley S (1998) Advanced backcross QTL analysis in tomato. I. Identification of QTLs for traits of agronomic importance from *Lycopersicon hirsutum*. *Theor Appl Genet* 97:381–397
- Bernardo R (2008) Molecular markers and selection for complex traits in plants: learning from the last 20 years. *Crop Sci* 48:1649–1664
- Bernardo R, Charcosset A (2006) Usefulness of gene information in marker-assisted recurrent selection: a simulation appraisal. *Crop Sci* 46:614–621
- Bernier J, Kumar A, Serraj R, Spaner D, Atlin G (2008) Review: Breeding upland rice for drought resistance. *J Sci Food Agric* 88:927–939
- Beumer KJ, Trautman JK, Bozas A, Liu JL, Rutter J, Gall JG, Carroll D (2008) Efficient gene targeting in *Drosophila* by direct embryo injection with zinc-finger nucleases. *Proc Natl Acad Sci USA* 105:19821–19826
- Bhushan D, Pandey A, Choudhary MK, Datta A, Chakraborty S, Chakraborty N (2007) Comparative proteomics analysis of differentially expressed proteins in chickpea extracellular matrix during dehydration stress. *Mol Cell Proteomics* 6:1868–1884
- Bibikova M, Carroll D, Segal DJ, Trautman JK, Smith J et al (2001) Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol Cell Biol* 21:289–297
- Bibikova M, Golic M, Golic KG, Carroll D (2002) Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. *Genetics* 161:1169–1175
- Bibikova M, Beumer K, Trautman JK, Carroll D (2003) Enhancing gene targeting with designed zinc finger nucleases. *Science* 300:764
- Bishop GJ, Harrison K, Jones JDG (1996) The tomato Dwarf gene isolated by heterologous transposon tagging encodes the first member of a new cytochrome P450 family. *Plant Cell* 8:959–969
- Bogdanove AJ, Voytas DF (2011) TAS effectors: customizable proteins for DNA targeting. *Science* 333:1843–1846
- Bozas A, Beumer KJ, Trautman JK, Carroll D (2009) Genetic analysis of zinc-finger nuclease-induced gene targeting in *Drosophila*. *Genetics* 182:641–651
- Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH et al (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol* 18:630–634
- Brent R (2000) Genomic biology. *Cell* 100:169–183
- Brown PJ, Upadyayula N, Mahone GS, Tian F, Bradbury PJ, et al (2011) Distinct genetic architectures for male and female inflorescence traits of maize. *PLoS Genet* 7:e1002383
- Buckler E (2007) TASSEL: Trait analysis by association, evolution, and linkage. User manual. <http://www.maizegenetics.net/tassel>
- Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ et al (2009) The genetic architecture of maize flowering time. *Science* 325:714–718
- Burley SK, Almo SC, Bonanno JB, Capel M, Chance MR et al (1999) Structural genomics: beyond the Human Genome Project. *Nat Genet* 23:151–157
- Burton RA, Gibeau DM, Bacic A, Findlay K, Roberts K et al (2000) Virus-induced silencing of a plant cellulose synthase gene. *J Plant Cell* 12:691–705
- Cai CQ, Doyon Y, Ainley WM, Miller JC, Dekelver RC et al (2009) Targeted transgene integration in plant cells using designed zinc finger nucleases. *Plant Mol Biol* 69:699–709

- Casa AM, Mitchell SE, Hamblin MT, Sun H, Bowers JE et al (2005) Diversity and selection in sorghum: simultaneous analyses using simple sequence repeats. *Theor Appl Genet* 111:23–30
- Cavanagh C, Morell M, Mackay I, Powell W (2008) From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Curr Opin Plant Biol* 11:215–221
- Chain P, Kurtz S, Ohlebusch S, Slezak T (2003) An applications-focused review of comparative genomics tools: capabilities, limitations and future challenges. *Brief Bioinform* 4:106–123
- Chang WW, Huang L, Shen M, Webster C, Burlingame AL, Roberts JK (2000) Patterns of protein synthesis and tolerance of anoxia in root tips of maize seedlings acclimated to a low-oxygen environment and identification of proteins by mass spectrometry. *Plant Physiol* 122:295–318
- Chapman MA, Pashley CH, Wenzler J, Hvala J, Tang S, Knapp SJ, Burke JM (2008) A genomic scan for selection reveals candidates for genes involved in the evolution of cultivated sunflower (*Helianthus annuus*). *Plant Cell* 20:2931–2945
- Christopher M, Mace E, Jordan D, Rodgers D, McGowan P et al (2007) Applications of pedigree-based genome mapping in wheat and barley breeding programs. *Euphytica* 154:307–316
- Ciais P, Reichstein M, Viovy N, Granier A, Ogee J et al (2005) Europe-wide reduction in primary productivity caused by the heat and drought in 2003. *Nature* 437:529–533
- Clark RM, Schweikert G, Toomajian C, Ossowski S, Zeller G et al (2007) Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science* 317:338–342
- Clayton TA, Lindon JC, Cloarec O, Antti H, Charuel C et al (2006) Pharmaco-metabonomic phenotyping and personalized drug treatment. *Nature* 440:1073–1077
- Clegg MT, Allard RW, Kahler AL (1972) Is the gene the unit of selection? Evidence from two experimental plant populations. *Proc Natl Acad Sci USA* 69:2474–2478
- Corona G, Rizzolio F, Giordano A, Toffoli G (2012) Pharmaco-metabolomics: an emerging “Omics” tool for the personalization of anticancer treatments and identification of new valuable therapeutic targets. *J Cell Phys* 227:2827–2831. doi:10.1002/jcp. 24003
- Crossa J, Burgueno J, Dreisigacker S, Vargas M et al (2007) Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. *Genetics* 177:1889–1913
- CSIRO (2011) A MAGIC approach to improve wheat quality. [http://www.csiro.au/en/Organisation-Structure/Divisions/Plant-Industry/MAGIC\\_FFF\\_ResearchOverview.aspx](http://www.csiro.au/en/Organisation-Structure/Divisions/Plant-Industry/MAGIC_FFF_ResearchOverview.aspx)
- de Koning DJ, Haley CS (2005) Genetical genomics in humans and model organisms. *Trends Genet* 21:377–381
- Dixon J, Braun HJ, Kosina P, Couch J (eds) (2009) Wheat facts and futures. CIMMYT, Mexico
- Doyon Y, McCammon JM, Miller JC, Faraji F, Ngo C (2008) Heritable targeted gene disruption in zebra fish using designed zinc-finger nucleases. *Nat Biotechnol* 26:702–708
- Dressman D, Yan H, Traverso G, Kinzler KW, Vogelstein B (2003) Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc Natl Acad Sci USA* 100:8817–8822
- Earley EJ, Jones CD (2011) Next-generation mapping of complex traits with phenotype-based selection and introgression. *Genetics* 189:1203–1209
- Eathington SR, Crosbie TM, Edwards MD, Reiter RS, Bull JK (2007) Molecular markers in a commercial breeding program. *Crop Sci* 47(S3):S154–S163
- Edward M, Johnson L (1994) RFLPs for rapid recurrent selection. In: Analysis of molecular marker data. Joint plant breeding symposium series. Crop Science Society of America, Madison, WI, pp 33–40
- Ehrenreich IM, Torabi N, Jia Y, Kent J, Martis S et al (2010) Dissection of genetically complex traits with extremely large pools of yeast segregants. *Nature* 464:1039–1042
- Eisen JA (1998) Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. *Genome Res* 8:163–167
- Ellis JG, Lawrence GJ, Luck JE, Dodds PN (1999) Identification of regions in alleles of the flax rust resistance gene *L* that determine differences in gene-for-gene specificity. *Plant Cell* 11:495–506

- Eshed Y, Zamir D (1995) An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield associated QTL. *Genetics* 141:1147–1162
- Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 8:186–194
- Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8:175–185
- FAO (2011) Food and agriculture organization of the United Nations, State of food insecurity in the world 2010–2011. FAO, Rome
- FAOSTAT (2009) <http://faostat.fao.org/default.aspx>
- Fedurco M, Romieu A, Williams S, Lawrence I, Turcatti G (2006) BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Res* 34:e22
- Fehr WR (1984) Genetic contributions to yield gains of five major crop plants (Spl Pub No 7). Crop Science Society of America, Madison, WI
- Fiehn O, Kloska K, Altmann T (2001) Integrated studies on plant biology using multiparallel techniques. *Curr Opin Biotechnol* 12:82–86
- Fisher RA, Edmeades GD (2010) Breeding and cereal yield progress. *Crop Sci* 50:585–598
- Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. *Annu Rev Plant Biol* 54:357–374
- Foley JE, Yeh JR, Maeder ML, Reyon D, Sander JD, Peterson RT, Joung JK (2009) Rapid mutation of endogenous zebrafish genes using zinc finger nucleases made by Oligomerized Pool Engineering (OPEN). *PLoS One* 4:e4348
- Gale MD, Devos KM (1998) Plant comparative genetics after 10 years. *Science* 282:656–658
- Geurts AM, Cost GJ, Freyvert Y, Zeitler B, Miller JC et al (2009) Knock out rats via embryo microinjection of zinc-finger nucleases. *Science* 325:433–435
- Giles J (2007) Key biology databases go wiki. *Nature* 445:691
- Godfray H CJ, Beddington JR, Crute IR, Haddad L, Lawrence D et al (2010) Food security: the challenge of feeding 9 billion people. *Science* 327:812–818
- Goldberg AD, Banaszynski LA, Noh KM, Lewis PW, Elsaesser SJ et al (2010) Distinct factors controls histone variant H3.3 localization at specific genomic regions. *Cell* 140:678–691
- González-Martínez SC, Wheeler NC, Ersoz E, Nelson CD, Neale DB (2007) Association genetics in *Pinus taeda* L. I. wood property traits. *Genetics* 175:399–409
- Grandillo S, Tanksley SD (2003) Advanced backcross QTL analysis: results and perspectives. In: Tuberosa R, Phillips RL, Gale M (eds) Proceedings of the international congress on in the wake of the double helix: from the green revolution to the gene revolution, 27–31 May 2003, Bologna, Italy. Avenue Media, Italy, pp 115–132
- Grant MR, McDowell JM, Sharpe AG, de Torres Zabala M, Lydiate DJ, Dangl JL (1998) Independent deletions of a pathogen-resistance gene in *Brassica* and *Arabidopsis*. *Proc Natl Acad Sci USA* 95:15843–15848
- Grossman A, Takahashi H (2001) Macronutrient utilization by photosynthetic eukaryotes and the fabric of interactions. *Annu Rev Plant Physiol Plant Mol Biol* 52:163–210
- Gur A, Zamir D (2004) Unused natural variation can lift yield barriers in plant breeding. *PLoS Biol* 2:e245
- Hajjar R, Hodgkin T (2007) The use of wild relatives in crop improvement: a survey of development over the last 20 years. *Euphytica* 156:1–13
- Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* 69:315–324
- Hamblin MT, Buckler ES, Jannink JL (2011) Population genetics of genomics-based crop improvement methods. *Trends Genet* 27:98–106
- Harlan HV, Martini ML (1929) A composite hybrid mixture. *J Am Soc Agron* 487–490
- Hedrick PW (1987) Gametic disequilibrium: proceed with caution. *Genetics* 117:331–341

- Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. *Theor Appl Genet* 38:226–231
- Holte S, Quiaioit F, Hsu L, Davidov O, Zhao LP (1997) A population based family study of a common oligogenic disease. Part I: Association/aggregation analysis. *Genet Epidemiol* 14:803–807
- Horst I, Welham T, Kelly S, Kaneko T, Sato S, Tabata S, Parniske M, Wang TL (2007) TILLING mutants of *Lotus japonicus* reveal that nitrogen assimilation and fixation can occur in the absence of nodule-enhanced sucrose synthase. *Plant Physiol* 144:806–820
- Hospital F, Moreau L, Lacoudre F, Charcosset A, Gallais A (1997) More on the efficiency of marker-assisted selection. *Theor Appl Genet* 95:1181–1189
- Howes NK, Woods SM, Towley Smith TF (1998) Simulations and practical problems of applying multiple marker assisted selection and doubled haploids to wheat breeding programs. *Euphytica* 100:225–230
- Huang X, Feng Q, Qian Q, Zhao Q, Wang L et al (2009) High throughput genotyping by whole-genotyping resequencing. *Genome Res* 19:1068–1076
- Huang X, Sang T, Zhao Q, Wei X, Feng Q et al (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet* 42:961–967
- Hunkapiller T, Kaiser RJ, Koop BF, Hood L (1991) Large-scale and automated DNA sequence determination. *Science* 254:59–67
- Issaq HJ, Conrads TP, Janini GM, Veenstra TD (2002) Methods for fractionation, separation and profiling of proteins and peptides. *Electrophoresis* 23:3048–3061
- Jackson SP, Bartek J (2009) The DNA-damage response in human biology and disease. *Nature* 461:1071–1078
- Jansen RC, Nap JP (2001) Genetical genomics: the added value from segregation. *Trends Genet* 17:388–391
- Johal GS, Briggs SP (1992) Reductase activity encoded by the *HM1* disease resistance gene in maize. *Science* 258:985–987
- Johal GS, Balint-Kurti P, Weil CF (2008) Mining and harnessing natural variation: a little MAGIC. *Crop Sci* 48:2066–2073
- Johnson R (2004) Marker-assisted selection. *Plant Breed Rev* 24:293–309
- Karayiorougou M, Sobin C, Blundell ML, Galke BL, Malinova L, Goldberg P, Ott J, Gogos JA (1999) Family-based association studies support a sexually dimorphic effect of COMT and MAOA on genetic susceptibility to obsessive-compulsive disorder: extending the transmission disequilibrium test (TDT) to examine genetic heterogeneity. *Biol Psychiatry* 45:1178–1189
- Kazuki S (2006) Gene identification through metabolomics in plants. *Cell Technol* 25:1399–1403
- Kearsey MJ, Hynes V (1994) QTL analysis: a simple ‘marker regression’ approach. *Theor Appl Genet* 89:698–702
- Keurentjes JJ, Angenent GC, Dicke M, Dos Santos VA, Molenaar J et al (2011) Redefining plant systems biology: from cell to ecosystem. *Trends Plant Sci* 6:183–190
- Kim YG, Cha J, Chandrasegaran S (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci USA* 93(1156):1160
- Kosina P, Reynolds MP, Dixon J, Joshi AK (2007) Stakeholder perception of wheat production constraints, capacity building needs, and research partnerships in developing countries. *Euphytica* 157:475–483
- Koumproglou R, Wilkes TM, Towson P, Wang XY, Beyon J et al (2002) STAIS: a new genetic resource for functional genomics studies of *Arabidopsis*. *Plant J* 31:355–364
- Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, et al (2009) A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genet* 5: e1000551
- Kraakman ATW, Niks RE, Van den Berg PM, Stam P, Van Eeuwijk FA (2004) Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivar. *Genetics* 168:435

- Krysan PJ, Young JC, Tax F, Sussman MR (1996) Identification of transferred DNA insertions within *Arabidopsis* genes involved in signal transduction and ion transport. *Proc Natl Acad Sci USA* 93:8145–8150
- Kumagi MH, Donson J, Della-Cioppa G, Harvey D, Hanley K, Grill LK (1995) Cytoplasmic inhibition of carotenoid biosynthesis with virus-derived RNA. *Proc Natl Acad Sci USA* 92:1679–1683
- Kumar GR, Sakthivel K, Sundaram RM, Neeraja CN, Balachandran SM et al (2010) Allele mining in crops: prospects and potentials. *Biotechnol Adv* 28:451–461
- Kump KL, Bradbury PJ, Buckler ES, Belcher AR, Oropeza-Rosas M et al (2010) Genome-wide association study of quantitative resistance to Southern leaf blight in the maize nested association mapping population. *Nat Genet* 43:163–168
- Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124:743–756
- Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121:185–199
- Lashkari DA, DeRisi JL, McCusker JH, Namath AF, Gentile C et al (1997) Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc Natl Acad Sci USA* 94:13057–13062
- Latha R, Rubia L, Bennett J, Swaminathan MS (2004) Allele mining for stress tolerance genes in *Oryza* species and related germplasm. *Mol Biotechnol* 27:101–108
- Lewontin RC (1964) The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* 49:49–67
- Lieber MR (2010) The mechanism of double-strand DNA break repair by the non-homologous DNA end-joining pathway. *Annu Rev Biochem* 79:181–211
- Liester D, Kurth J, Laurie DA, Yano M, Sasaki T, Graner A, Schulz-Lefert P (1998) Rapid reorganization of resistance gene homologues in cereal genomes. *Proc Natl Acad Sci USA* 95:370–375
- Lincoln SE, Daly MJ, Lander ES (1993) Mapping genes controlling quantitative traits using MAPMAKET/QTL version 1.1: a tutorial and reference manual. Whitehead Institute, Cambridge, MA
- Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Harvey Millar A, Ecker JR (2008) Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* 133:523–536
- Lobell DB, Burke MB, Tebaldi C, Mastrandrea MD, Falcon WP, Naylor RL (2008) Prioritizing climate change adaptation needs for food security in 2030. *Science* 319:607–610
- Louder O, Chaillou S, Camilleri C, Bouchez D, Daniel-Vedele F (2002) Bay-0x Shahdara recombinant inbred line population: a powerful tool for the genetic dissection of complex traits in *Arabidopsis*. *Theor Appl Genet* 104:1173–1184
- Macdonald SJ, Long AD (2007) Joint estimates of quantitative trait locus effect and frequency using synthetic recombinant populations of *Drosophila melanogaster*. *Genetics* 176:1261–1281
- Maeder ML, Thibodeau-Beganny S, Osiak A, Wright DA, Anthony RM et al (2008) Rapid 'open-source' engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol Cell* 31:294–301
- Marteinssen RA (1998) Functional genomics: probing plant gene function and expression with transposons. *Proc Natl Acad Sci USA* 95:2021–2026
- Maxted N, Scholten MA, Codd R, Ford-Lloyd BV (2007) Creation and use of a national inventory of crop wild relatives. *Biol Conserv* 140:142–159
- Mayer JE, Pfeiffer WH, Beyer P (2008) Biofortified crops to alleviate micronutrient malnutrition. *Curr Opin Plant Biol* 11:166
- McCallum CM, Comai L, Greene EA, Henikoff S (2000) Targeted screening for induced mutations. *Nat Biotechnol* 18:455–457

- Meng X, Noyes MB, Zhu LJ, Lawson ND, Wolfe SA (2008) Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat Biotechnol* 26:695–701
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829
- Meyer RC, Steinfath M, Lisee J, Becher M, Witucka-Wall H et al (2007) The metabolic signature related to high plant growth rate in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 104:4759–4764
- Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW et al (1999) Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding super-family. *Plant J* 20:317–332
- Michelmore R (2000) Genomic approaches to plant disease resistance. *Curr Opin Plant Biol* 3:126–131
- Mitra RD, Church GM (1999) In situ localized amplification and contact replication of many individual DNA molecules. *Nucleic Acids Res* 27:e34
- Mitra RD, Shendure J, Olejnik J, Edyta Krzymanska O, Church GM (2003) Fluorescent *in situ* sequencing on polymerase colonies. *Anal Biochem* 320:55–65
- Mizoi J, Nakamura M, Nishida I (2006) Defects in CTP: phosphorylethanolamine cytidyl transferase affect embryonic and postembryonic development in *Arabidopsis*. *Plant Cell* 18:3370–3385
- Moehle EA, Rock JM, Lee YL, Jouvenot Y, DeKolver RC et al (2007) Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proc Natl Acad Sci USA* 104:3055–3060
- Moller IS, Gilliam M, Jha D, Mayo GM, Roy SJ et al (2009) Shoot Na<sup>+</sup> exclusion and increased salinity tolerance engineered by cell type-specific alteration of NA<sup>+</sup> transport in *Arabidopsis*. *Plant Cell* 21:2163–2178
- Morbitz R, Romer P, Boch J, Lahaye T (2010) Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. *Proc Natl Acad Sci USA* 107:21617–21622
- Morrell PL, Buckler ES, Ross-Ibarra J (2012) Crop genomics: advances and applications. *Nat Rev Genet* 12:85–96
- Moynahan ME, Jasin M (2010) Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. *Nat Rev Mol Cell Biol* 11:196–207
- Myles S, Peiffera J, Brown PJ, Ersoza ES, Zhang ZW, Costicha DE, Buckler ES (2009) Association mapping: critical considerations shift from genotyping to experimental design. *Plant Cell* 21:2194–2202
- Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320:1344–1349
- Nelson JC (1997) QGENE: Software for marker-based genomic analysis and breeding. *Mol Breed* 3:239–245
- Nelson DE, Repetti PP, Adams TR, Creelman RA, Wu JR et al (2007) Plant nuclear factor Y (NF-Y) B sub-units confer drought tolerance and lead to improved corn yields on water-limited acres. *Proc Natl Acad Sci USA* 104:16450
- Nordborg M, Tavaré S (2002) Linkage disequilibrium: what history has to tell us. *Trends Genet* 18:83–90
- Oliver SG, Winson MK, Kell DB, Baganz R (1998) Systematic functional analysis of the yeast genome. *Trends Biotechnol* 16:373–378
- Ortiz R, Sayre K, Govaerts B, Gupta R, Subbarao G et al (2008) Climate change: can wheat beat the heat? *Agric Ecosyst Environ* 126:46–58
- Pan Q, Wendel J, Fluhr R (2000) Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. *J Mol Evol* 50:203–213

- Pandey A, Chakraborty S, Datta A, Chakraborty N (2008) Proteomics approach to identify dehydration responsive nuclear proteins from chickpea (*Cicer arietinum* L.). *Mol Cell Proteomics* 7:88–107
- Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J et al (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457:548–551
- Paul Voosen (2009) Molecular breeding makes crops hardier and more nutritious. *Scientific American*. <http://www.scientificamerican.com/article.cfm?id=molecular-breeding-crops-genetics-rice-soy-corn-wheat>
- Peltonen-Sainio P, Jauhiainen L, Trnka M, Olesen JE, Calanca P et al (2010) Coincidence of variation in yield and climate in Europe. *Agric Ecosyst Environ* 139:482–489
- Prada D (2009) Molecular population genetics and agronomic alleles in seed banks: searching for a needle in a haystack? *J Exp Bot* 60:2541–2552
- Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000) Association mapping structured populations. *Am J Hum Genet* 67:170–181
- Quarrie SA, Gulli M, Calestani C, Steed A, Marmioli N (1994) Location of a gene regulating drought-induced abscisic acid production on the long arm of chromosome 5A of wheat. *Theor Appl Genet* 89:794–800
- Rangan L, Constantino S, Khush GS, Swaminathan MS, Bennett J (1999) The feasibility of PCR-based allele mining for stress tolerance genes in rice and related germplasm. *Rice Genet Newsl* 16:43–47
- Ratcliff F, Martin-Hernandez AM, Baulcombe DC (2001) Technical Advance. Tobacco rattle virus as a vector for analysis of gene function by silencing. *Plant J* 25:237–245
- Ravi K, Vadez V, Isobe S, Mir RR, Guo Y et al (2011) Identification of several small-effect main QTL and large number of epistatic QTLs for drought tolerance in groundnut (*Arachis hypogaea* L.). *Theor Appl Genet* 122:1119–1132
- Reynolds M, Bonnett D, Chapman Scott C, Furbank Rober T, Manès Y et al (2011) Raising yield potential of wheat. I. Overview of a consortium approach and breeding strategies. *J Exp Bot* 62 (2):439–452
- Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. *Science* 273:1516–1517
- Rivkin MI, Vallejos CE, McClean PE (1999) Disease-resistance related sequences in common bean. *Genome* 42:41–47
- Rockman MV, Kruglyak L (2008) Breeding designs for recombinant inbred advanced intercross lines. *Genetics* 179:1069–1078
- Roessner U, Beckles DM (2009) Metabolite measurements. In: Junker B, Schwender J (eds) *Plant metabolic networks*. Springer, Heidelberg, pp 39–69
- Saghatelian A, Tauger SA, Want EJ, Hawkins EG, Siuzdak G, Cravatt BF (2004) Assignment of endogenous substrates to enzymes by global metabolite. *Biochemistry* 16:14332–14339
- Sanchez AC, Subudhi PK, Rosenow DT, Nguyen HT (2002) Mapping QTLs associated with drought resistance in sorghum (*Sorghum bicolor* L. Moench). *Plant Mol Biol* 48:713–726
- Sanger F, Nicklen S, Coulson R (1977) DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci USA* 74:5463–5467
- Santiago Y, Chan E, Liu PQ, Orlando S, Zhang L et al (2008) Targeted gene knockout in mammalian cells using engineered zinc finger nucleases. *Proc Natl Acad Sci USA* 105:5809–5814
- Saranga Y, Menz M, Jiang CX, Wright RJ, Yakir D, Paterson AH (2001) Genomic dissection of genotype × environment interactions conferring adaptation of cotton to arid conditions. *Genome Res* 11:1988–1995
- Sari-Gorla M, Krajewski P, Di Fonzo N, Villa M, Fropa C (1999) Genetic analysis of drought tolerance in maize by molecular markers. II. Plant height and flowering. *Theor Appl Genet* 99:289–295
- Sax K (1923) The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* 8:552–560



- Scarcelli N, Cheverud JM, Schaal BA, Kover PX (2007) Antagonistic pleiotropic effects reduce the potential adaptive value of the *FRIGIDA* locus. *Proc Natl Acad Sci USA* 104:16986–16991
- Schauer N, Semel Y, Roessner U, Gur A, Balbo I et al (2006) Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nat Biotechnol* 24:447–454
- Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270:467–470
- Shendure J, Ji H (2008) Next-generation DNA sequencing. *Nat Biotechnol* 26:1135–1145
- Shendure J, Porreca GJ, Reppas NB, Lin XX, McCutcheon JP et al (2005) Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* 309:1728–1732
- Shukla VK, Doyon Y, Miller JC, DeKolver RC, Moehle EA et al (2009) Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. *Nature* 459:437–441
- Slade AJ, Fuerstenberg SI, Loeffler D, Steine MN, Facciotti D (2005) A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. *Nat Biotechnol* 23:75–81
- Stahl EA, Dwyer G, Mauricio R, Kreitman M, Bergelson J (1999) Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. *Nature* 400:667–671
- Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. *Plant J* 3:739–744
- Suneson CA (1956) An evolutionary plant breeding method. *Agron J* 48:188–191
- Talukdar A, Zhang GQ (2007) Construction and characterization of 3-S Lines, an alternative population for mapping studies in rice. *Euphytica* 156:237–246
- Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* 277:418–423
- Tanksley SD, Nelson JC (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from un-adapted germplasm into elite breeding lines. *Theor Appl Genet* 92:191–203
- Taylor J, King RD, Altmann T, Fiehn O (2002) Application of metabolomics to plant genotype discrimination using statistics and machine learning. *Bioinformatics* 18:241–248
- Tester M, Langridge P (2010) Breeding technologies to increase crop production in a changing world. *Science* 327:818–822
- Teulat B, Monneveux P, Wery J, Borriès C, Souyris I, Charrier A, This D (1997) Relationships between relative water content and growth parameters in barley: a QTL study. *New Phytol* 137:99–107
- Tian F, Bradbury PJ, Brown PJ, Hung H et al (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat Genet* 43:159–162
- Till BJ, Reynolds SH, Weil C, Springer N, Burtner C et al (2004) Discovery of induced point mutations in maize genes by TILLING. *BMC Plant Biol* 4:12
- Townsend JA, Wright DA, Winfrey RJ, Fu FL, Maeder ML, Joung JK, Voytas DF (2009) High frequency modification of plant genes using engineered zinc-finger nucleases. *Nature* 459:442–445
- Tsiridis E, Giannoudis PV (2006) Transcriptomics and proteomics: advancing the understanding of genetic basis of fracture healing. *Injury* 37S:S13–S19
- Tsugita A, Kamo M, Kawakami T, Ohki Y (1996) Two-dimensional electrophoresis of plant proteins and standardization of gel patterns. *Electrophoresis* 17:855–865
- Turcatti G, Romieu A, Fedurco M, Tairi AP (2008) A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. *Nucleic Acids Res* 36:e25
- Ungerer MC, Halldorsdottir SS, Modiszewski JL, Mackay TFC, Purugganan M (2002) Quantitative trait loci for inflorescence development in *Arabidopsis thaliana*. *Genetics* 160:1133–1151
- Urnov FD, Miller JC, Lee YL, Beausejour CM, Rock JM et al (2005) Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* 435:646–651
- Utz HF, Melchinger AE (1996) PLABQTL: a program for composite interval mapping of QTL. *J Quant Trait Loci* 2:1–5

- Vigouroux Y, McMullen M, Hittinger CT, Houchins K, Schulz L et al (2002) Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. *Proc Natl Acad Sci USA* 99:9650–9655
- Villas-Bôas SG, Roessner U, Hansen M, Smedsgaard J, Nielsen J (2007) Metabolome analysis: an introduction. Wiley, Hoboken, NJ
- Wang H, Hanash S (2003) Multi-dimensional liquid phase based separations in proteomics. *J Chromatogr B Analyt Technol Biomed Life Sci* 787:11–18
- Wang XY, Wan JM, Su CC, Wang CM, Shen WB et al (2004) QTL detection for eating quality of cooked rice in a population of chromosome segment substitution lines. *Theor Appl Genet* 110:71–79
- Wang J, Chapman SC, Bonnett DG, Rebetzke GJ, Crouch J (2007) Application of population genetic theory and simulation models to efficiently pyramid multiple genes via marker-assisted selection. *Crop Sci* 47:582–588
- Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10:57–63
- Wang J, Roe B, Macmil S, Yu Q, Murray JE et al (2010) Micro-collinearity between autopolyploid sugarcane and diploid sorghum genomes. *BMC Genomics* 11:261
- Weinstein JN (1998) Fishing expeditions. *Science* 282:628–629
- Weinthal D, Tovkach A, Zeevi V, Tzfira T (2010) Genome editing in plant cells by zinc finger nucleases. *Trans Plant Sci* 15:308–321
- Whitham S, Dinesh-Kumar SP, Choi D, Hehl R, Corr C, Baker B (1994) The product of the tobacco mosaic virus resistance gene *N*: similarity to toll and the interleukin-1 receptor. *Cell* 78:1101–1115
- Whittaker JC, Thompson R, Visscher PM (1996) On the mapping of QTL by regression of phenotypes on marker type. *Heredity* 77:23–32
- Wilkins MR, Ou K, Appel RD, Sanchez JC et al (1996) Rapid protein identification using N-terminal “sequence tag” and amino acid analysis. *Biochem Biophys Res Commun* 221 (3):609–613
- Wilson IW, Schiff CL, Hughes DE, Somerville SC (2001) Quantitative trait loci analysis of powdery mildew disease resistance in the *Arabidopsis thaliana* accession Kashmir-1. *Genetics* 158:1301–1309
- Winkler RG, Frank MR, Galbraith DW, Feyereisen R, Feldmann KA (1998) Systematic reverse genetics of transfer-DNA-tagged lines of *Arabidopsis*. Isolation of mutations in the cytochrome P450 gene super family. *Plant Physiol* 118:743–750
- Wong CK, Bernardo R (2008) Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations. *Theor Appl Genet* 116:815
- Wood AJ, Te-Wen L, Bryan Z, Catherine SP, Edward JR et al (2011) Targeted genome editing across species using ZFNs and TALENs. *Science* 333:307
- Wright AJ, Mowers RP (1994) Multiple regression for molecular-marker, quantitative trait data from large F<sub>2</sub> populations. *Theor Appl Genet* 89:305–312
- Xiao J, Li J, Yuan L, Tanksley SD (1996) Identification of QTLs affecting traits of agronomic importance in a recombinant inbred population derived from a sub-specific rice cross. *Theor Appl Genet* 92:230–244
- Yadav RS, Sehgal D, Vadez V (2011) Using genetic mapping and genomics approaches in understanding and improving drought tolerance in pearl millet. *J Exp Bot* 62:397–408
- Yamasaki M, Tenaillon MI, Vroh Bi I, Schroeder SG, Sanchez-Villeda H et al (2005) A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. *Plant Cell* 17:2859–2872
- Yan YP, Moulton J (2005) Protein family clustering for structural genomics. *J Mol Biol* 353:744–759
- Yu J, Buckler ES (2006) Genetic association mapping and genome organization of maize. *Curr Opin Biotechnol* 17:155–160
- Yu J, Pressoir G, Briggs WH, Vroh BI, Yamasaki M et al (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* 38:203–208

- Yu JM, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. *Genetics* 178:539–551
- Zeng ZB (1994) Precision mapping of quantitative trait loci. *Genetics* 136:1457–1468
- Zhu CS, Gore M, Buckler ES, Yu JM (2007) Status and prospects of association mapping in plants. *Plant Genome* 1:5–20