

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Progression of DNA Marker and the Next Generation of Crop Development

Herry S. Utomo, Ida Wenefrida and Steve D. Linscombe
Louisiana State University Agricultural Center
 USA

1. Introduction

Advancement in genomic technology has been the main thrust for the progression of DNA markers that is now approaching a critical point in providing a platform for the next generation of varietal development. Improving total yield production to meet the increasing need to feed the world population remains the major goal. However, achieving more sophistication in providing high quality crop products to meet the emerging demand for better nutritional values and food functionalities will increasingly become important goals. Progression in high throughput marker analyses, significant reduction in the cost per data point, sophistication in computational tools, and creation of customized sets of markers for specific breeding applications are continuing and expected to have direct implications for highly efficient crop development in the near future. An advanced DNA marker system can be used to accomplish breeding goals, as well as achieve various scientific goals. The goals encompass a wide array of targets from understanding the function of specific genes so detailed that the quality of gene output or products can be controlled to attaining a global view of genomic utility to improve crop development efficiency. The combination of molecular understanding at the individual gene levels and genetic manipulation at the genome levels may lead to a significant yield leap to meet global food challenges.

Historically, plant breeding always integrates the latest innovations to enhance crop improvement. Starting out with the prehistoric selection based on systematic visual observations leading to the first plant domestication (Harlan, 1992), crop development was further enhanced by employing Darwin's scientific principles of hybridization and selection, then applying Mendel's principles of association between genotype and phenotype, and now through DNA markers and genomics that will lead to the next generation of crop development. Recent progression in high throughput marker genotyping, genome scanning, sequencing and re-sequencing, molecular breeding and bio-informatics, software and algorithm, and precise phenotyping (Delseny et al. 2010; Edwards & Batley, 2010; Varshney et al., 2009; Mochida & Shinozak, 2010; Davey et al., 2011) is a conduit for the next generation of crop development that uses different views and avenues to approach the same goal. This chapter will discuss various aspects of DNA molecular markers associated with crop development. They include the progression of molecular marker technology, prospect and current limitations, applications in unraveling global genetic potential, and specific utilization for exploiting global genetic sources and re-purpose some of the traits to fulfill local demand in the given environmental conditions.

2. Advancement in DNA marker technology

2.1 DNA marker

Genetic markers are heritable polymorphisms found among individuals or populations that can be measured (Davey et al., 2011). They become the center point of modern genetics to answers many important questions in population genetics, ecological genetics and evolution. A genetic marker is a gene or DNA sequence with a known location on a chromosome that can be used to identify cells, individuals, and species, or traits of interest. It can be described as an observable variation from mutation or alteration in the genomic loci. A genetic marker can be a short DNA sequence, such as a sequence surrounding a single base-pair change (single nucleotide polymorphism, SNP), or a long one, like microsatellite or simple sequence repeats (SSRs). DNA markers are available in many forms. In addition to older types of markers, such as RFLP, RAPD, and AFLP, combination of molecular markers, simple sequence repeats (SSRs), indels, PCR-based SNPs, and CAPS (cleaved amplified polymorphic sites) markers, are markers that can be used for many purposes including genotyping.

DNA markers can be used to identify or verify a true identity of cultivars or breeding lines, the F₁ hybrids, seed purity, or intra-varietal variation. The analyses are relatively simpler, easier, and more accurate than phenotypic evaluation. Few well-selected markers are typically adequate for providing definitive discrimination of cultivars in question. One of the major marker contributions is better understanding of genetic diversity, population structures, genetic relationship among subspecies, genetic relationship within specific germplasm collections, and family relationship among breeding lines and cultivars. Molecular markers provide high quality genetic data that may not be possible to produce through other genetic methods. Genetic relationship and connectivity among available germplasm is invaluable knowledge to realize the overall genetic potential that can be managed to develop more productive cultivars. Gene survey to find specific target alleles from different groups of germplasm is very useful for a breeder to efficiently use and manage available germplasm. An example in rice includes the survey of *Pi-ta* blast-resistant genes among a large collection of rice lines (Wang, et al., 2007b).

2.2 Simple sequence repeats (SSRs)

SSR markers are co-dominant, reliable, randomly distributed throughout the genome, and highly polymorphic. Due to its superiority, SSRs have been used widely for mapping purposes. They are generally transferable between mapping populations. SSRs consist of tandemly repeated short nucleotide units 1–6 bp in length. The abundance of SSR markers in the genome is generated through SSR mutational rates that occur between 4×10^{-4} to 5×10^{-6} per allele and per generation (Primmer et al., 1996; Vigouroux et al., 2005). The mutation is caused predominantly through 'slipped strand mis-pairing' during DNA synthesis that will result in the gain or loss of one or more repeat units (Levinson and Gutman, 1987). The most widely distributed SSRs are di-, tri- and tetranucleotide repeated motifs, such as (CA)_n, (AAT)_n and (GATA)_n (Tautz & Renz, 1984). Prior to whole genome sequence, SSR markers initially were developed from expressed sequence tags (ESTs) and bacterial artificial chromosome (BAC) end sequences. For example, prior to the completion of rice genome sequence, 2,240 SSR markers were identified using publicly available BAC and PAC clones (McCouch et al., 2002). The completion of whole genome sequence in many

plant species has led to the identification of numerous SSR markers. A large number of SSR markers totaling 797,863 SSRs were identified in three monocots *Brachypodium*, sorghum and rice and three dicots *Arabidopsis*, *Medicago* and *Populus* using their whole genome sequence information (Sonah et al., 2011). Mono-nucleotide repeats were the most abundant repeats, and the frequency of repeats decreased with increase in motif length both in monocots and dicots. The frequency of SSRs was higher in dicots than in monocots both for nuclear and chloroplast genomes. Based on SSR analyses of these six species, GC-rich repeats were the dominant repeats in monocots, with the majority of them being present in the coding region that involved in different biological processes, predominantly binding activities. Their locations on the physical map can be accessed through various online databases. The SSR markers are available to the public providing marker density of approximately 51 SSR per Mb or less and are suitable for use in mapping and various applications of MAS.

2.3 Single nucleotide polymorphism (SNP)

Single nucleotide polymorphism (SNP) is an individual nucleotide base difference between two DNA sequences. SNP markers are the ultimate form of genetic polymorphism and, therefore, may become the predominant markers in the future. SNPs can be categorized according to nucleotide substitution, i.e. transitions (C/T or G/A) or transversions (C/G, A/T, C/A or T/G). In both human and plant, C/T transitions constitute 67% of the SNPs observed (Edwards et al., 2007). In human, about 90% of variation is attributed to SNPs, equating to approximately 1 SNP in every 100–300 bases. Based on partial genomic sequence information from barley, soybean, sugarbeet, maize, cassava, potato and other crops, typical SNP frequencies are also in the range of one SNP every 100–300 bp (Edwards et al., 2007; Hyten et al., 2010). Rice serves a crop model and, therefore, discovery and utilization of SNP markers in rice can be enhanced and perfected for other crop plants. A genome-wide analysis in rice cultivars Nipponbare (japonica subspecies) and 93-11 (indica subspecies) revealed 1,703,176 SNPs and 479,406 indels (Shen et al., 2004), giving a frequency of approximately 1 SNP per 268 bp. Using alignments of the improved whole-genome shotgun sequences for japonica and indica rice, the SNP frequencies vary from 3 SNPs/kb in coding sequences to 27.6 SNPs/kb in the transposable elements (Yu et al., 2005). Major re-sequencing effort based on hybridization based re-sequencing of 20 diverse *O. sativa* varieties by Perlegen BioSciences for the OryzaSNP project, a total of 159,879 SNPs were identified (McNally et al. 2009). The quality of the SNPs in this study was very good, but the SNP discovery pool covered only about 100Mb of the genome and had a low discovery rate (approximately 11% within the tiled 100 Mb region). To improve the SNP discovery pool, hundreds of diverse *O. sativa*, *O. rufipogon*/*O. nivara*, *O. glaberrima* and *O. barthii* accessions are currently being re-sequenced by groups in several countries (McCouch et al., 2010).

SNP detections can be done through a number of methods, including gel electrophoresis, fluorescence resonance energy transfer (FRET), fluorescence polarization, arrays or chips, luminescence, mass spectrophotometry, and chromatography. Fluorescence detection method currently is the most widely used in high-throughput genotyping. Fluorescence has been used in different detection applications, including plate readers, capillary electrophoresis and DNA arrays. Many types of fluorescent plate readers are available with the capability to detect fluorescence in a 96- or 384-well format by using a light source and narrow band-pass filters to select the excitation and emission wavelengths and enable semi-

quantitative steady state fluorescence intensity readings to be made (Jenkins & Gibson, 2002). It also has been applied for genotyping with TaqMan, Invader and rolling-circle amplification. Fluorescence plate readers allow measurement of additional fluorescence parameters, including polarization, lifetime and time-resolved fluorescence, and fluorescence resonance energy transfer. In addition, mass spectrometry and light detection are also used for high throughput SNP genotyping.

DNA chip or gene chip is a SNP detection platform for high-throughput genotyping. It consists of a collection of microscopic DNA spots attached to a solid surface. This is one of the fastest research developing areas. More than 1.8 million markers (about 906,600 SNPs and 946,000 probes) are available from the Affymetrix® Genome-Wide Human SNP Array 6.0 for the detection of copy number variation. Luminex has developed a panel of 100 bead sets with unique fluorescent labels that can be processed by flow analyzer. Besides detecting SNPs, genotyping, or re-sequencing mutant genomes, DNA microarrays has been used to measure gene expression. SNP detection also can be done using mass spectrometry based on molecular weight difference of DNA bases. Variation of this technique includes MALDI-TOF (matrix-assisted laser desorption/ionization-time of flight) mass spectrometry that uses allele-specific incorporation of two alternative nucleotides into an oligonucleotide probe to allow measurement of the mass of the extended primers. This approach can also detect PEX products in multiplex very efficiently. Both DNA microarrays developed by Affymetrix (Santa Clara, USA) and a high-density biochip assay by Illumina Inc. (San Diego, USA) are two major chip-based high-throughput genotyping systems that offer different levels of multiplexes of several thousands (Yan et al., 2010). When an ultra-high density SNP map is used, QTL gain detection efficiency has improved considerably compared to using maps from traditional RFLP/SSR markers (Yu et al., 2011).

2.4 DArT (diversity array technology) and RAD (restriction site associated DNA)

Dramatic advancement of SSR and SNP marker technology and their applications have been achieved in important organisms, including humans and a number of model animals and crops. However, discovering sequence polymorphism in non-model species, especially 'orphan' crop and other crops that have complex, polyploid genomes, remains slow. DArT (diversity arrays technology) is a microarray hybridization-based marker system that can be used to overcome the problem, since it does not require prior knowledge of genetic or genomic sequence (Yang et al., 2011; Alves-Freitas et al., 2011; Jaccoud et al., 2001; Wenzl et al., 2004). It has relevant applications for species with complex genomes and especially for the 'orphan' crops important for Third World countries. In addition to its high throughputness, DArT is relatively quick, highly reproducible, and cost effective about tenfold lower than SSR markers per data point (Xia et al., 2005). It is designed for open use and not covered by exclusive patent rights. Users can freely specify the scope of genetic analyses and it can be expanded as needed.

Typical DArT analyses include 1) Constructing a reference library representing the genetic diversity of a species through extraction of total genomic DNA (metagenome) from a pool of individuals (i.e. a group of cultivated genotypes or to be combined together with their wild relatives, followed by complexity reduction to produce genomic representation, and cloning using suitable vector and *E. coli*; 2) Preparing "discovery array" containing individual clones; 3) Generating genomic representations of individual lines studied; 4) Hybridizing

with genomic representations of all genomes in the metagenome library; 5) Identifying polymorphic clones and assembling polymorphic data into “genotyping array”; and 6) Genotyping analyses, including construction of linkage mapping or other type of analyses.

Beside its great potentials, DArT has inherent limitations. First, DArT markers are dominant markers (present or absent), which restrict its value in some applications. Second, it is a microarray-based technique that involves several steps, including preparation of genomic representation for the target species, cloning, and data management and analysis. These steps require expertise, additional cost, and also utilization of supporting software, such as DArTsoft, DArTdb, and DArtsoft 7. These may pose some limitation to its full utilization potential in the developing countries. Beside a slow start centered around the team that developed the system, an increasing number of independent research groups now have routinely utilized the methodology involving a broader range of species for various purposes, including linkage mapping (Yang et al., 2011), genotyping of closely related species (Alves-Frietas et al., 2011), genotyping very large and complex genomes such as wheat (Paux et al., 2008) and sugarcane (Wei et al., 2010).

More recently, a variety of microarrays (including tiling/cDNA/oligonucleotide arrays) also has been used to develop the so-called RAD markers for study of genomewide variations associated with restriction sites for individual restriction enzymes. For this purpose, first a genome-wide library of RAD tags is developed from genomic DNA, which is then used for hybridization on to the chosen microarray to detect all restriction site-associated variations in a single assay. The development of RAD tags involves the following steps: (i) digestion of genomic DNA with a specific restriction enzyme; (ii) ligation of biotinylated linkers to the digested DNA; (iii) random shearing of ligated DNA into fragments smaller than the average distance between restriction sites, leaving small fragments with restriction sites attached to the biotinylated linkers; (iv) immobilization of these fragments on streptavidin-coated beads; and (v) release of DNA tags from the beads by digestion at the original restriction sites. This process specifically isolates DNA tags directly flanking the restriction sites of a particular restriction enzyme throughout the genome. The RAD tags from each of a number of samples, when hybridized on to a microarray, allows high-throughput identification and/or typing of differential hybridization patterns. These markers have clear advantage over the existing marker systems (for example, restriction fragment length polymorphisms, AFLPs and DArT markers) that could assay only a subset of SNPs that disrupt restriction sites. RAD markers were successfully developed in a number of organisms, including fruit fly, zebrafish, threespine stickleback, and *Neurospora* (Lewis et al., 2007; Miller et al., 2007a, b) and will certainly find their way in most of the laboratories working on higher plants.

Another high throughput restriction-based marker is RAD (Restriction site Associated DNA) markers that can be used for genetic mapping. To generate RAD markers, RAD tags (the DNA sequences immediately flanking each instance of a particular restriction enzyme site throughout the genome) need to be isolated. This involves digesting DNA with a particular restriction enzyme, ligating biotinylated adapters to the overhangs, randomly shearing the DNA into much smaller fragments than the average distance between restriction sites, and isolating the biotinylated fragments using streptavidin beads (Miller et al., 2007b). Different RAD tag densities can be obtained by utilizing different restriction enzymes during the isolation process. Once RAD tags are isolated, they can be used for microarray analysis (Miller et al., 2007a; Lewis et al., 2007).

As an alternative, RAD analyses can be incorporated into high-throughput sequencing (i.e. on the Illumina platform; Baird et al., 2008). For that, the RAD tag isolation procedure will need to be modified. After the production of DNA fragments much smaller than the average distance between restriction sites by random shearing, it will be followed by preparation of the sheared ends and ligation of the second adapter, and amplification of specific fragments that contain both adapters using PCR. The first adapter contains a short DNA sequence barcode. Different DNA samples can be prepared with different barcodes to allow for sample tracking when multiple samples are sequenced in the same reaction (Hohenlohe et al., 2010; Baird et al., 2008). These RAD tags can then be subjected to high-throughput sequencing for more efficient RAD mapping. The sequencing approach produces higher genetic marker density than microarray methods.

2.5 Random, genic, and functional markers

DNA markers can be classified as 1) random markers (anonymous or neutral markers) when they are derived at random from polymorphic sites across the genome, 2) gene targeted or candidate gene markers when they are derived from polymorphisms within genes, and 3) functional markers when they are derived from polymorphic sites within genes that are causally associated with phenotypic trait variation (Andersen & Lübberstedt, 2003; Wei et al., 2009). Each marker type may be used for specific purposes. Random markers, for example, can be used as an effective tool for establishing a breeding system, studying a gene flow among natural populations, and determining a genetic structure of population or characterizing a GeneBank collection (Xu et al., 2005). Although the predictive value of a random marker depends on the known linkage phase between marker and target locus alleles (Lübberstedt et al., 1998), so far, a random marker remains the marker system of choice for marker-assisted breeding and QTL analyses in a wide variety of crop plants (Semagn et al., 2010; Xu, 2003b).

Both genic and functional markers are derived within the genes. Therefore, they are correlated well with gene function and have a high predictive value for the targeted gene in selection (Anderson & Lübberstedt, 2003; Wei et al., 2009). Because of that, they are most suited for use in marker-assisted breeding. The number of both genic and functional markers increases substantially in the recent years due to the availability of DNA sequence information from whole genome sequence projects that are available publically for a number of plant species, including rice, soybean, cassava, maize, barley, wheat, potato, and tomato (Mochida & Shinozak, 2010). Sequence data of fully characterized genes and full-length cDNA clones are also available for some plant species, including those described above. The sequence data for ESTs, genes, and cDNA clones can be downloaded from GeneBank and scanned for identification of markers, including SSRs which are typically referred to as EST-SSRs or genic microsatellites. Many gene-derived SSR markers for maize, for example, have been developed from genes using the available information in GeneBank and their primer sequences are available at www.maizeGDB.org.

Genic SSRs are more transferable across species than genomic markers, especially when the primers are designed from more conserved coding regions (Varshney et al., 2005). EST-SSR markers could, therefore, be used in related species where information on SSRs or ESTs is limited. These markers can also be used effectively for comparative mapping (Shirasawa et al., 2011; Yu et al., 2004; Varshney et al., 2005; Oliveira et al., 2009). EST-SSRs can be used to

produce high quality markers, but they are often less polymorphic than genomic SSRs (Wang et al., 2011; Aggarwal et al., 2007; Eujayl et al., 2002; Thiel et al., 2003). The EST resources can be further mined for SNPs (Ramchiary et al., 2011; Li et al., 2009).

2.6 Physical and molecular genetic map

2.6.1 Physical map

A physical map provides information on the order of genetic components on the chromosomes in terms of physical distance units (base pairs). Deciphering actual biological functions of the physical map hold the key to unravel the overall genetic potential of organisms. However, a construction of a whole-genome physical map is crucial. It provides a solid blueprint for quantifying species evolution, revealing species-specific features, delineating ancestral biological functions shared by a certain group of plant species, predicting and interpreting regulatory signatures, and for practical purposes identifying candidate genes needed in crop improvement through sequences of functional or structural orthologs among closely related or model species. The construction of a physical map has been the critical component in numerous genome projects, including the Human Genome Project (HGP), the first genome project initiated in 1990 and completed in 13 years, to produce and integrate genetic, physical, gene and sequence maps. As of today, 25 published plant genome sequence (complete, publicly available, and can be used without restriction) is available, including for potato, grape, *Arabidopsis thaliana*, *A. lyrata*, *Theellungiella*, *Brassica rapa*, poplar, cucumber, cannabis, apple, strawberry, soybean, Pigeon pea, lotus, Medicago, Date palm, maize, sorghum, *Brachypodium*, rice, *selaginella*, and *Physcomitrella* (CoGePedia, 2011). Rice (*Oryza sativa*) genome sequence was the second (after *Arabidopsis*) to be published in plants, but it is the first monocot, grass, grain, and food crop genome. Its original published genome published in 2002 is consisted of a dual publication from two independent groups, using two subspecies of rice, japonica and indica. The current version of the rice genome contains ~370 megabases of sequence and 40,577 non-transposon related genes spread across 12 chromosomes (CoGePedia, 2011).

Physical mapping can be carried out using BAC-by-BAC or clone-by-clone strategy using two-step progression; First is the establishment of BAC clones (typically 100–150 kb) for the target genome/chromosome together with a set of overlapping clones representing a minimal tiling path (MTP) to be ordered along the chromosomes of the target genome. Shotgun sequencing is then applied to the individually mapped clones of the MTP. The DNA from each BAC clone is randomly fragmented into smaller pieces to be cloned into a plasmid and then subjected to Sanger sequencing (dideoxy sequencing or chain termination method) or sequenced directly using Next Generation Sequencing (NGS) technologies. The resulting sequence data are then aligned so that identical sequences overlap and contiguous sequences (contigs) are assembled into a finished sequence. Unlike the Sanger sequencing technology, NGS technologies are based on massive parallel sequencing, do not require bacterial cloning, and only rely on the amplification of single isolated DNA molecules. Tens of millions of single-stranded DNA molecules can be immobilized on a solid surface, such as a glass slide or on beads, and analyze them in a massively parallel way providing extremely rapid sequencing. Physical mapping can also be done through whole-genome shotgun (WGS) strategy involving the assembly of sequence reads generated in a random, genome-wide fashion. The entire target genome (chromosome) is fragmented into pieces of certain

sizes that can either be subcloned into plasmid vectors or sequenced directly using NGS technologies. Highly redundant sequence coverage across the genome or chromosome can be generated through sequence reads from many subclones and using various computational methods, the sequences are assembled to produce a consensus sequence.

One of the most expected outcomes of the genome sequence is high-throughput development of molecular markers to assist genetic analysis, gene discovery and breeding programs (Fukuoka et al. 2010). Because of the genome sequence, rice, for example, is now rich in tools for mapping and breeding. It has high density SSRs of about 51 SSR per Mb, comprehensive SNPs (1,703,176 SNPs, approximately one SNP every 268 bp), insertion-deletion polymorphisms (IDPs) and custom designed (candidate gene) markers for marker-assisted breeding (Feuillet et al, 2011). Upon the completion of genome sequence, various efforts have been dedicated to tributary SNP markers identified from the sequence into breeder's chips where it can be used as a breeding tool (McCouch et al. 2010). A combination of low-, medium- and high resolution SNP assays are being developed for variety of purposes. The low density SNP chips, the 384-SNP OPAs are particularly attractive to the breeding and geneticists because of their reliability and require little technical adjustment once they are designed and optimized. Hundreds or thousands of individuals can be assayed within a short time window and are relatively inexpensive compared to the time, labor and bioinformatics requirements of other marker technologies.

2.6.2 Genetic map

A genetic map is produced by counting recombinant phenotypes revealing important genetic layouts of the organism. Marker based-genetic linkage map is generally constructed using the same principles for constructing classical genetic maps. The components of mapping include selection of markers, development of mapping populations from selected parental lines, genotyping and phenotyping each individual in the mapping population using molecular markers; and constructing linkage maps from the phenotypic and marker data. To define a recombination frequency between two linked genetic markers, genetic distance units known as centiMorgans (cM) or map units are used. Two markers are 1 cM apart if they are found to be separated in one of 100 progeny. However, 1 cM does not always correspond to the same length of DNA physical distance. The actual length of DNA per cM is referred to as the physical to genetic distance. In the genome areas where recombination occurs frequently (recombination hot spots), shorter length of DNA per cM - as low as 200 kb/cM, can be found. The characteristic of recombination hot spot is that the gene or genes where crossovers occurred are mostly located in very small genetic intervals, consisting mostly of 1-2 genes, and that those genes almost always harbor one or more single feature polymorphisms (Singer et al., 2006). In other parts where recombination may be suppressed, the physical to genetic distance can be 1500 kb per 1 cM. The lowest recombination rates typically occur at the centromeres due to heavily methylated heterochromatin (Haupt et al., 2001). The proportion of recombinant gametes depends on the rate of crossover during meiosis and is known as the recombination frequency (r). The maximum proportion of recombinant gametes is 50% when crossover between two genetic loci has occurred in all the cells. This is equivalent to non-linked genes where the two loci are inherited independently. The recombination frequency depends on the rate of crossovers which in turn depends on the linear distance between two genetic loci. Recombination

frequencies, range from 0 (complete linkage) to 0.5 (complete independent inheritance). A measure of the likelihood that genes are linked is expressed as the logarithm of the odds (LOD). The LOD score (logarithm (base 10) of odds), developed by Newton E. Morton, is a statistical test to determine that two loci are linked. Positive LOD scores indicate the presence of linkage, whereas negative LOD scores indicate that linkage is less likely. By convention, a LOD score greater than 3.0 is considered evidence for linkage.

2.7 Mapping populations

There are various types of populations that can be used to create genetic maps, develop marker linked to target genes, and facilitate marker verification. The most common populations created for mapping purposes include F_2 s, backcrosses (BCs), double haploids (DHs), recombinant inbred lines (RILs), and near isogenic lines (NILs). In association mapping, natural populations are used. DHs are produced from chromosome doubling of haploids via in vivo and in vitro. They have several advantages over other diploid populations of F_2 s, F_3 s, or BCs, since no dominance or dominance-related epistasis effects involves in the genetic model. As a result, additive, additive-related epistasis, and linkage effects can be investigated properly. As a permanent population, DH lines can be replicated as many times as desired across different environments, seasons and laboratories, providing endless genetic material for phenotyping and genotyping and to evaluate the genotype-by-environment interaction (Forster & Thomas, 2004; Bordes et al., 2006). In DH populations, the additive component of genetic variance is larger than that of F_2 and BC populations. Detailed quantitative genetics associated with DH populations have been previously discussed, including detection of epistasis, estimation of genetic variance components, linkage test, estimation of gene numbers, genetic mapping of polygenes and tests of genetic models and hypotheses (Choo et al., 1985; Bordes et al., 2006).

Recombinant inbred lines or random inbred lines (RILs) can be produced through various inbreeding procedures. They include full-sib mating for open-pollinated plants and selfing for self-pollinated plants. In self-pollinated plants, RIL can be developed through a bulking method where hybrids are bulk planted and harvested until F_5 to F_8 before they are planted by families. RIL can also be produced through single seed descent (SSD) where one or several seeds are harvested from each F_2 plant and planted to produce the next generation until F_5 to F_8 . Near-isogenic lines are the product of inbreeding through successive backcrossing.

2.8 Mapping software and tools

Almost all molecular maps on the first generation of molecular markers, such as RFLPs, were constructed using MAPMAKER/EXP (Table 1). For severe distortion of segregation, statistical modifications will be needed and MAPDISTO can be used to solve this problem. JOINMAP can be used for construction genetic linkage for BC_1 , F_2 , RIL, F_1 - and F_2 -derived DH and out-breeder full-sib families. It can combine ('join') data derived from several sources into an integrated map, with several other functions, including linkage group determination, automatic phase determination for out-breeder full-sib family, several diagnostics and map charts (van Ooijen & Voorrips, 2001). A software package CMAP, a web-based tool, allows users to view comparisons of genetic and physical maps. The

package also includes tools for curating map data (Ware et al., 2002). There are many commercial or freely available software packages for establishing association between marker genotypes and trait phenotypes. The most commonly used are QTL CARTOGRAPHER, MAPQTL, PLABQTL and QGENE. All of these only handle bi-allelic populations, while MCQTL (Jourjon et al., 2005) can perform QTL mapping in multi-allelic situations, including bi-parental populations from segregating parents, or sets of biparental, bi-allelic populations. The most frequently used QTL software during the 1980s and 1990s was MAPMAKER/QTL. MAPL allows a user to get results on segregation ratio, linkage test, recombination value, group markers, and order of markers by metric multi-dimensional scaling, and to draw a QTL map through interval mapping and analysis of variance (ANOVA).

A currently widely used QTL mapping software is QTL CARTOGRAPHER (Table 1). PLABQTL uses composite interval mapping with many functions similar to QTL CARTOGRAPHER. QTL can be localized and characterized in populations derived from a biparental cross by selfing or production of DHs. Simple and composite interval mapping are performed using a fast multiple regression procedure and can be used for QTL $\times$ environment interaction analysis (Utz & Melchinger, 1996). Recently, QGENE has been rewritten in the Java language and can be used for analyses of trait and QTL permutation and simulation for populations and as well as traits. Several software packages can be used for constructing linkage maps in out-crossing plant species, using full-sib families derived from two outbreed (non-inbreeding) parent plants (Garcia et al., 2006). Bayesian QTL mapping has received a lot of attention in recent years. Several software packages have been developed; For example, BQTL can perform maximum likelihood estimation of multi-gene models, Bayesian estimation of multi-gene models using Laplace Approximations, and interval and composite interval mapping of genetic loci. BLADE was for Bayesian analysis of haplotypes for LD mapping. MULTIMAPPER is a Bayesian QTL mapping software for analyzing backcross, DH and F_2 data from designed crossing experiments of inbred lines (Martinez et al., 2005). MULTIMAPPER/OUTBRED for populations derived from out-bred lines. Several mapping software packages were developed for QTL mapping for some specific situations. MCQTL was developed for simultaneous QTL mapping in multiple crosses and populations (Jourjon et al., 2005), including diallel cross modeling of the QTL effects using multiple related families. MAPPOP was developed for selective and bin mapping by selecting samples from mapping populations and for locating new markers on pre-existing maps (Vision et al., 2000). In addition, QTLNETWORK was developed for mapping and visualizing the genetic architecture underlying complex traits for experimental populations from a cross between two inbred lines (Yang et al., 2008).

Web-based QTL analytical tools are also available. Some of the tools developed in other system can potentially serve as a model for plants. WEBQTL (Table 1) provides dense error-checked genetic maps, as well as extensive gene expression data sets (Affymetrix) acquired across more than 35 strains of mice. To map QTLs in out-bred populations, QTL EXPRESS (Seaton et al., 2002) was developed for line crosses, half-sib families, nuclear families and sib-pairs. It provides two options for QTL significance tests: permutation tests to determine empirical significance levels and bootstrapping to estimate empirical confidence intervals of QTL locations.

Association or LD mapping is another mapping tool using unstructured populations of unrelated individuals, germplasm accessions, or randomly selected cultivars. Prior to LD mapping, genotype units are subjected to statistical analysis to remove population structure, which can cause false positive associations due to circumstantial correlations rather than real linkage. To meet the requirement, the STRUCTURE software (Pritchard et al., 2000) can be used. Some software packages have already included the population structure analysis functionality. STRAT, as a companion program to STRUCTURE, uses a structured association method for LD mapping, enabling valid case-control studies even in the presence of population structure (Pritchard et al., 2000). TASSEL can be used for trait analysis by association, evolution and linkage, which performs a variety of genetic analyses including LD mapping, diversity estimation and LD calculation (Zhang, et al., 2006). MIDAS can be used for analysis and visualization of inter-allelic disequilibrium between multi-allelic markers (Gaunt et al., 2006). With PEDGENIE, any size pedigree may be incorporated into this tool, from independent individuals to large pedigrees and independent individuals and families may be analyzed together. GENERECON is another software package for LD mapping using coalescent theory. It is based on a Bayesian Markovchain Monte Carlo method for fine-scale LD mapping using high-density marker maps. Genome-wide association (GWA) studies are used to find the link between genetic variations and common diseases in humans, as well as agronomic traits in plants. A well-powered GWA study will involve the measurement of hundreds of thousands of SNPs in thousands of individuals. Statistical tools developed for GWA studies include GENOMIZER, MAPBUILDER, CATS (Table 1).

3. Application and contributions of DNA markers to cultivar development

3.1 DNA marker utilizations

One of the most successful practical uses of molecular markers to date is gene introgression and pyramiding. Publicly available information on gene-marker association for a number of important agronomic traits can readily be used to introgress and pyramid these genes into elite breeding lines used in cultivar development. Marker-assisted backcrossing (MABC) is a straight forward method to introgress or move target gene(s) from parental donors to parental recipients. It involves successive backcrossing to remove the genetic background of the donor while recovering genetic properties of recurrent parents as much as possible. Statistical methods and schedule of backcrosses to create effective MABC have been reviewed in various papers (Hospital, 2001; Hospital & Charcosset, 1997; Herzog & Frisch, 2011). MABC with marker-based genome scanning has allowed a speedy recovery of most recurrent genome in a few crosses (Frisch et al., 1999; Frisch & Melchinger, 2005). MABC can also be used to develop cleaner near isogenic lines by minimizing carried over donor segments flanking the target locus, providing precise introgression of individual genes for detailed characterization of the QTLs. Marker-assisted gene pyramiding has been successfully utilized to combine multiple genes of male sterility (Nas et al., 2005) or to provide broader-spectrum of resistance against major diseases, such as rice blast and bacterial blight (Yoshimura et al., 1996; Jeung et al., 2006). Individual genes have unique reactions against pathogenic races and some of them have overlapping spectra that make selection based on disease reactions or symptoms more challenging. This problem can easily be overcome using molecular markers linked to individual disease-resistant genes allowing effective selection to be carried out to stack the genes.

No.	Name	Common Use
1	MAPMAKER/EXP	The first and most frequently used mapping software for map construction in the early era of DNA markers developed by the Whitehead Institute (Lander et al., 1987).
2	MAP MANAGER CLASSIC	Provides a graphical presentation and interactive tool to map Mendelian loci for codominant markers, using backcrosses or RILs in plants or animals (Manly, 1993).
3	MAP MANAGER CLASSIC	Provides a graphical presentation and interactive tool to map Mendelian loci for codominant markers, using backcrosses or RILs in plants or animals (Manly, 1993).
4	MAPDISTO	Can be used to address segregation distortion in segregating populations, such as backcross, double haploid (DH) and RIL populations. It computes and draws genetic maps through a graphical interface and analyzes marker data by showing segregation distortion due to differential viability of gametes or zygotes. Maps or data from multiple populations derived from different crosses can be combined into single or consensus maps through joint mapping. (ftp: http://mapdisto.free.fr/)
5	JOINMAP	For construction of genetic linkage maps for several types of mapping populations. (www.kyazma.nl/index.php/mc.JoinMap/).
6	CMAF	Provides comparative function developed as a web-based tool to allow users to view comparisons of genetic and physical maps.
7	MAPMAKER/QTL;	A sister software package to MAPMAKER/EXP, developed by Lander et al. (1987) based on maximum likelihood estimation of linkage between marker and phenotype using interval mapping to deal with simple QTLs and several standard populations
8	MAPL	Allows a user to get results on segregation ratio, linkage test, recombination value, group markers, and order of markers by metric multi-dimensional scaling, and to draw a QTL map through interval mapping and analysis of variance (ANOVA). Developed by Ukai et al., 1995),
9	QTL CARTOGRAPHER	Implements several statistical methods using multiple markers simultaneously, including composite interval and multiple composite interval mapping. Interaction between identified QTLs can also be estimated.
10	PLABQTL	Uses composite interval mapping.
11	QGENE	Intended for comparative analyses of QTL mapping data sets, developed in 1991 as a map and population simulation program, to which QTL analyses were added later on (www.qgene.org/)

No.	Name	Common Use
12	MAPQTL	Calculates QTL positions on genetic maps for several types of mapping populations, including BC1s, F2s, RILs, DHs. It can also be used for QTL interval mapping, composite interval mapping and non-parametric mapping using functions for automatic cofactor selection and permutation test.
13	BQTL	Used for the mapping of genetic traits from line crosses and RILs (Borevitz et al., 2002).
14	BLADE	Used for Bayesian analysis of haplotypes for LD mapping (Liu et al., 2001; Lu, et al., 2003).
15	MULTIMAPPER/ OUTBRED	Has multi uses, including populations derived from out-bred lines.
16	WEBQTL	Used for exploring the genetic modulation of thousands of phenotypes gathered over a 30-year period by hundreds of investigators using reference panels of recombinant inbred strains of mice in web-based applications.
17	TASSEL	Comprehensive LD-based QTL mapping for trait analysis by association, evolution and linkage, which performs a variety of genetic analyses, including LD mapping, diversity estimation and LD calculation (Zhang et al., 2006).
18	MIDAS	For analysis and visualization of inter-allelic disequilibrium between multi-allelic markers (Gaunt et al., 2006)
19	PEDGENIE	Used as a general purpose tool to analyze association and transmission disequilibrium (TDT) between genetic markers and traits in families of arbitrary size and structure (Allen-Brady et al., 2006)
20	GENOMIZER	A platform independent Java program for the analysis of GWA experiments.
21	PLINK	A whole genome LD analysis toolset (Purcell et al., 2007),
22	MAPBUILDER	For chromosome-wide LD mapping (Abad-Grau et al., 2006)
23	CATS	Calculates the power and other useful quantities for two-stage GWA studies (Skol et al., 2006)

Table 1. List of QTL mapping software

3.2 QTL mapping

A long history of breeding suggests that grain yield is controlled by many genes with small effects. For this type of trait, applicability of finding and introgressing QTLs are limited since estimates of QTL effects for minor QTLs are often inconsistent. Even though these minor QTLs could show consistent effects, pyramiding these minor QTLs is increasingly challenging as the number of QTLs pyramided into one line increases (Bernardo, 2008). Inconsistency of estimated QTL effects for complex traits controlled by many minor genes brings the following important consequences. Due to limited transferability of estimated QTL effects across different populations for traits such as grain yield, QTL mapping will have to be repeated for each breeding population. Under this condition, Marker-assisted recurrent selection (MARS) is suitable since genotyping, phenotyping, and construction of

selection index are repeated for each population (Koeber, 2003; Campbell et al., 2003). Because GXE interactions have a great influence on complex traits controlled by many QTLs with minor effects, QTL mapping from the same population needs to be conducted in each target set of environments. Finally, because the effects of sampling errors are high, population size of 500 to 1,000 is suggested (Beavis, 1994).

Mapping of multiple trait complexes in multiple environments can be conducted by employing algorithmic models to predict the association of genetic markers with trait of interest based on the effect of variance and covariance of the analyzed. These models allow the designing of new mapping frameworks and simulation tools, and the association to be extrapolated to the progeny of the plant or genetic materials tested in multiple environments. There is no limitation on the number of environments where the traits are scored. Based on simulation studies (Howes et al., 1998; Wang et al., 2007a), combining favorable marker alleles for more than 12 unlinked QTLs appears to be not feasible. The breeder may initially target a large number of QTLs but expects to accept having fewer QTL alleles fixed in a recombinant inbred. Since the improvement can only be targeted in a limited QTL number, breeders need a high level of confidence that the target QTLs do not represent a false positive that implies stringent levels of significance, $P \leq 0.0001$, when identifying the QTLs initially. A stringent significant level, however, can lead to an upward bias in estimating QTL effects (Beavis, 1994; Xu, 2003a) and therefore lead to overly optimistic expectation of response from MAS. Based on empirical and simulation studies, selection responses are increased when less stringent significant levels of $P = 0.20$ to 0.40 were applied in MARS. These relaxed significant levels allow QTLs with smaller effects to be selected and these minor QTLs can exceedingly compensate for the higher frequency of false positive. Less stringent significant levels are acceptable for pointing QTL locations, and when the goal is to predict genotypic performance such as in MARS, more stringent significant levels are required for combining favorable QTLs in recombinant inbred, introgression, and gene discovery. Along this line, QTLs should ideally be tagged by the markers inside the QTLs, or closely linked, or flanking the QTLs. Based on simulation studies in maize, the response of MARS in a population size of 144 plants was highest when about 128 markers are used (Bernardo & Charcosset, 2006), indicating that markers should be placed 10 to 15 cM apart and, therefore, denser markers are not necessary for predicting the performance.

Methods for using genetic markers, such as gene sequence diversity information, to improve plant breeding in developing cultivars by predicting the values of phenotypic traits based on genotypic, phenotypic, and optional family relationship information to identify marker-trait associations in the first population and used to predict the value of the phenotypic trait in the second or target population (Smith et al., 2005). These locally important traits are complex qualitative traits that are affected by many genes, the environment, and interaction between genes and environments. The next wave of QTL mapping should be targeted for locally important QTLs directly associated with cultivar development, including the matrix QTLs. It has been suggested that specific targets need to be clearly defined before embarking into the QTL mapping (Bernardo, 2008). In the context described above, yield potential and its components, quality traits, and local adaptation are among the most important QTL mapping targets. The architecture of genetic matrix of these complex traits could be dissected through QTL mapping to provide critical information on genomic regions and

fragment sizes that produce the effects and relative importance of additive and non-additive gene action (Fridman et al., 2004).

Multiple QTL Mapping (MQM Mapping) using haplotyped putative QTL alleles has been used as a simple approach for mapping QTLs in plant breeding population (Jansen & Beavis, 2001). It described a method for mapping a phenotypic trait to correspond to chromosomal location. Statistical methods to correlate pedigree with multiple markers (haplotype) are used to determine identical-by-descent (IBD) data to map the phenotypic traits. The statistical model, HAPLO-IM+, HAPLO-MQM, and HAPLO-MQM+ are used for mapping traits to determine a single gene or QTL. This invention provides an efficient method for mapping phenotypic traits in interrelated plant populations. The basic principle of this method is clustering of the original parental lines into groups on the basis of their haplotypes for multiple genetic markers is the basic principle of this method. The effect of a QTL on the phenotype is modeled per haplotype group instead of per family, allowing an examination of the effects of haplotype-allele across families. Simulations of realistic plant breeding schemes have shown a significant increase in the power of QTL detection. This approach offers new opportunities for mapping and exploitation of QTL in commercial breeding programs. In addition, selection can be performed at any stage of a breeding program, including among genetically distinct breeding populations as a preselection to increase the selection index and to drive up the frequency of favorable haplotypes among the breeding populations, among segregating progeny from breeding population to increase the frequency of favorable haplotypes for the purpose of developing cultivars, among segregating progeny from a breeding population to increase the frequency of the favorable haplotypes prior to QTL mapping within this breeding population, and among parental lines from different heterotic groups in hybrid crops to predict the performance potential of different hybrids.

The index values generated from haplotype window-trait association allows pre-selection, which is widely considered as the next generation of MAS, to further economize breeding by not only removing the need of required phenotypic evaluation but also enabling screening of inbred lines prior to making crosses. Breeders can initiate their programs by selecting a list of crosses and building a model based on haplotypes carried by each parental line in the cross. Selecting a model from cross to cross and inclusion of target genomic regions in the model will increase the complexity of the models. If it is not controlled, it will compromise the predictive ability and selection gain. For controlling the model's complexity, Automatic Model Picking (AMP) algorithm can be employed. The relative strength of each cross can be predicted using the Best Linear Unbiased Predictions (BLUP) approach, calculated on parental lines using phenotypic data. Once the final model is determined, the full gain of for each trait is calculated and the frequency-adjusted predicted gain can be obtained based on expected allele frequency. An additional optimization step can be included to either decrease or increase the importance of the secondary trait in the model based on frequency-adjusted predicted gain. This method provides haplotype information that allows the breeder to make informed breeding decisions based on genotype rather than phenotype into predictive breeding.

3.3 Channeling molecular information into new cultivar development

Successful marker breeding requires integration of molecular information into cultivar development programs. The development and testing of a QTL mapping population and

the development of near-isogenic lines (NIL) can take several years before the results can be utilized (Monforte & Tanksley, 2000; Chaib et al., 2006). Because of that by the time the QTL identified, the recurrent parent used in the population development is probably commercially obsolete. The purified QTLs still require to be reintroduced into a competitive germplasm for commercial use via a time-consuming backcross scheme. Frampton (2008) has proposed a direct integration of genomic technologies into commercial plant breeding by designing specific crossing schemes to allow the development of marker profiles, QTL mapping of major gene loci, and new cultivars to be advanced simultaneously. The method can be applied repeatedly to achieve complete integration. The breeding population is developed through an initial cross, followed by two backcrosses and self-pollination of BCF₁ plants. Molecular marker development consists of QTL identification using the means of BC₂F₂ family, gene fine mapping, and new marker development using bulk-segregant analysis. Therefore, the method provides simultaneous development of a breeding population with molecular marker development and gene mapping, and integration of molecular marker platform with the breeding platform.

4. The bottom line: Prospect and current limitations

Significant progress has been achieved in marker detection methodology in term of speed and cost. Current efforts by various consortiums supported by both public and private entities are underway to push the development of genomic tools to make them more economically and logistically feasible for the breeders. High resolution of marker assay covering all important information across the genomes, such as SNP chip sets being developed, will provide a tremendous asset to mobilize and assemble critical alleles that can improve crop production systems in a significant way. Over 1,200 reports of mapped QTLs are available through various publications in 12 major crop species (Bernardo, 2008; Xu & Crouch, 2008). Each typically reported an average of 3 to 5 QTLs for the trait studied (Bernardo, 2008; Eathington et al., 2007). This large volume of published molecular marker-trait associations will continue to grow as a result from the abundant amount of available markers, high density molecular assays, and development of sophisticated user-friendly computer software, and improved cost and technical efficiency in marker analyses. Despite a significant influx of reported marker-QTL trait associations to date, successful exploitation of available mapped QTLs remains low, indicating a lack of synchronization between the QTLs reported and actual breeding goals in the cultivar development. Successful integration of genomic tools in the cultivar development requires sufficient knowledge of breeding materials from molecular perspectives. This is essential for marker-based accumulation of favorable alleles and the ability to predict the effects of new QTLs assembled during cultivar development. Dissection of individual QTLs will lead to a better understanding of their interaction, discovery of hidden QTLs, and a new way to characterize and classify QTLs to facilitate a speedy assembly of critical genes needed to maximize the end products, such as grain production or other specific quality traits of economic significance. The genes critical for maintaining local adaptation and standard industrial and market quality are also among the most important QTLs.

At present, there is a mounting gap between available QTL mapping information and marker-based QTL applications in cultivar development. This vast mostly unexplored area presents a tremendous opportunity for both public researchers and private industry to tag

pivotal genes, including their interactions that play important roles in grain production. Because of the massive nature of the undertaking in both financial and human powers, a number of consortiums have been formed and operated to achieve greater common goals nearly impossible to achieve by any single lab. Major companies that have sufficient resources have shown an increased intensity in their effort in this arena, though only very limited information goes to the public.

Publicly available plant databases provide a large amount of genomic data for a wide range of plant species. This wealth of knowledge, however, has not yet found its way into mainstream plant breeding due to several reasons; first, there is no apparent connection between the primary information generated in plant genomics and real life breeding application. Second, databases storing various bits of supportive information (e.g. pedigree, genotype and phenotype) are usually stored in different places and managed by different groups of scientists. And third, there is a gap between breeders and molecular geneticists in perceiving their focus of interest, i.e. tools and interfaces for bioinformatic data focuses vs. organism level. Integration of the fragmented information, views, and priorities will be one of many challenges to overcome. Bioinformatics data typically consist of cDNA and genomic sequence data, genetic maps of mutants, DNA markers and maps, candidate genes and quantitative trait loci (QTL), physical maps based on chromosome breakpoints, gene expression data and libraries of large inserts of DNA such as bacterial artificial chromosomes and radiation hybrids. Information flows from molecular markers to genetic maps to sequences and to genes. However, the relationship between breeding (i.e. germplasm, pedigree and phenotype) and sequence-based information has not been established. An example of how genetic information can be integrated into plant breeding programs to produce cultivars from molecular variation using bioinformatics and what crop scientists might want from bioinformatics have been previously discussed (Mayes et al., 2005). How to best utilize all relevant genomic information efficiently and comprehensively, and harnessing the power of informatics to support molecular breeding is a challenge to modern plant breeding.

Processing speed and cost of DNA markers have improved substantially in the past several years, resulting in a significant reduction in processing time and cost per data point. This is one of the research areas where most rapid development occurs. However, its current application as a breeding tool has not reached its potential fully. Marker approach involves a separate line of research activities, and almost in all cases, it requires substantial upfront support. In addition to the cost of genotyping and phenotyping, it requires lab facilities and bioinformatics personnel to analyze complex data. This vast mostly unexplored area poses current limitations, but it could also present a tremendous opportunity for both public researchers and private industry to tag pivotal genes, including their interactions that play important roles in grain production, and to subsequently protect their invention.

4.1 Common platform and supporting tools

An appropriate experimental design and data analysis are a critical component for successful application of molecular breeding. Various models of data flowchart and analytical tools to funnel DNA marker data into cultivar development have been proposed. However, they lack of simple-to-use guidelines to allow breeders to confidently select the appropriate design and analysis. Communications between genomics scientists, geneticists,

bioinformaticians, and breeders are still limited hampering the development of truly integrated tools for applied molecular breeding design, integrated mapping, and MAS. Decision support tools for marker breeding that can model, simulate, and analyze most of the pre-existing genetic conditions will help breeders design and implement the efficient breeding scheme in term of cost and time using the optimum combination of MAS and phenotypic selection. Similarly important are decision support tools that include sample collection and depositing, retrieving, and tracking data, and also acquiring, collecting, processing, and mining databases.

For information-driven plant breeding, databases and supporting tools that allow an interchangeable flow of information through communicable platforms that required minimum maintenance and updates are critical. The use of universal language within different platforms will strengthen interaction among breeders, database curators, bioinformaticians, molecular biologists and tool developers. Interchangeable format and data content across all plant species are needed to develop a universal database. Current models, such as the one provided by Gene Ontology and Plant Ontology projects, offer a glimpse of future possibility in this area. An automatic ontological analysis has been used to develop biological interpretation of the data (Khatri et al., 2002). Currently, this approach becomes the standard for the secondary analysis of high-throughput experiments. A large number of tools have been developed for this purpose. Khatri and Draghici (2005) provided a review of detailed comparison for the 14 available tools using six different criteria; scope of the analysis, visualization capabilities, statistical model(s) used, correlation for multiple comparisons, reference microarray available, and installation issues and sources of annotation data. These analyses help researchers to select the most appropriate tool for a given type of analysis. Despite a few drawbacks in each tool associated with conceptual limitations of the current state-of-the-art in ontological analysis, this type of analysis has been generally adopted. These limitations are some of the challenges to overcome in order to create the next generation of secondary data analysis tools. Another major challenge is to construct a graphical presentation of systematic biological relationship that integrates gene, protein, metabolite, and phenotype data as suggested by Blanchard (2004). This will include an assembly of large-scale data sets into a more comprehensive presentation by minimizing high false positive rates and validating the existing models using probability and graph theory.

4.2 Added complexity in scope and time management

Crop development is a complex process, and molecular marker information further increases the complexity. To successfully apply molecular marker-assisted breeding, breeders have to structure their specific breeding methodologies to allow for the integration of empirical results from molecular marker analyses. All of molecular activities, including molecular analyses, establishment of genotypic-phenotypic associations, and molecular marker-based decision making, have to be completed in the same limited time frame in conventional breeding. They must be synchronized with seed planting preparation, progeny selection, yield trials, collecting phenotypic data, harvesting, data analyses, and use of off-season nurseries.

While some breeders have the access to computational infrastructure and statistical expertise needed to generate and analyze the gigabytes of genomic data, the majority will

depend on the availability of smaller subsets of genomic data that can be analyzed using an MS Excel spreadsheet. In the rice SNP system, for example, one of the current major efforts is to develop low-resolution SNP assay (through Affymetrix's custom-designed SNP genotyping arrays and Illumina's custom-designed SNP oligonucleotide pools assays (OPAs), or other platforms developed by KBiosciences) to address the problem (McCouch et al., 2010). In addition to reduce computation complexity, breeders will eventually be able to request targeted SNP detection assays that can be tailored into their specific breeding purposes or selecting their population base at a fraction the current cost of re-sequencing, particularly when the bioinformatic requirements are taken into account.

Breeders utilize breeding information from many different sources to obtain a description of genetic background and phenotypic traits under specific growing environment. The depth and types of information needed by individual breeders will vary greatly. However, data that critical for individual breeders will include some basic information, such as germplasm information (pedigree, genealogy, genetic stock data, etc.), genotypic information (DNA markers, sequences, and expression information), phenotypic data and environmental information. In addition, historical data preserved timely in the repository system can be used to reanalyze hypotheses and guide new research for molecular marker breeding. To obtain a high quality of mapping, both genotyping and phenotyping have to be conducted effectively. While molecular detection systems are rapidly enhanced, methods of phenotyping have not been improved as fast. Dissection of agronomically important QTLs requires phenotyping under target environments in multiple test sites. Proper techniques to ensure the consistency of phenotyping over multiple growing environments will need to be established.

5. Unraveling genetic potential globally

Providing sufficient food for an increasing world population is a tremendous challenge to overcome. Finding ways to boost the yield potential of major grain crops beyond current productivity levels, therefore, is critically important. One of the keys to solving the problem is to increase the ability to find novel alleles that are not present among cultivated species

Various studies show that wild species have a wider genetic diversity where critical alleles hidden or lost during the early domestication process and along the progression of modern breeding processes can be recovered. Extensive germplasm of various crop plants and their wild relatives are available in various places. In rice, for example, more than 102,547 accessions of Asian cultivated rice *O. sativa*, 1,651 accessions of African cultivated rice *O. glaberrima* and 4,508 accessions of wild ancestors are maintained in the International Rice Germplasm Collection (IRGC) at IRRI (McNally et al. 2006) in addition to an extensive rice germplasm collection in Japan, China, Taiwan, India, Korea, the USA, and many other countries. Relatives of rice species, such as *Oryza rufipogon* appear to have many new putative yield-related QTLs that can potentially be used to improve cultivated rice (Tan et al., 2007). Genome sequencing of wild species and map alignment are current ground breaking projects to provide a basic road to unravel the whole potential of wild species. The *Oryza* map Alignment Project (OMAP) is set to develop physical maps of 12 wild species to be aligned with the reference genome sequence of Nipponbare (Ammiraju et al., 2006; 2010; Wing et al., 2005; 2007). Sequence data will provide direct evidence of evolutionary path of *Oryza* genus. However, the most important expected outcomes from this current endeavor

are to find new genes and QTLs that can be used to improve grain production, levels of pest and disease tolerance, ability to tolerate stress and other less favorable growing environments. The ideas to unlock wild genetic variation to improve global grain production (McCouch et al., 2010; Fridman et al., 2004; Matsumoto et al., 2005) have, therefore, gained renewal interest from time to time. The precision of DNA markers to unravel the intercalating process of gene expression to determine the productivity of grain crop is needed to rediscover valuable alleles that can be funneled into the pipeline of cultivar development.

6. Practical utilization: Global source, local purpose

To survive in a very competitive market that demands high quality product, breeders have to assemble a series of genes that give rise to high yielding cultivars that have stable grain quality, disease resistance, optimum plant maturity and height, and are very adapted to target growing regions of a typically narrow niche of environments. These quality traits of industrial standards are critical for successful commercial production of crops in a modern era and often are the breeding priorities in current breeding programs. Long breeding selections have resulted in the formation of a specific matrix of complex QTLs that support quality traits required by the market. This matrix provides a skeleton for newer cultivars in grain crop breeding programs. Any efforts to improve current yield potential should, therefore, be built to correspondingly maintain or enhance the trait matrix. To stay competitive, breeders will be required to expand their crop to provide additional traits that are not currently available in their breeding populations. During the introgression of foreign traits into their breeding lines, all necessary matrix traits to produce high quality standards need to be maintained. Breeders have acquired detailed knowledge on the genetics underlying the matrix of these complex traits among individual breeding lines in the pipeline of cultivar development. Should molecular markers be employed in the breeding program, the same in-depth molecular knowledge must be acquired for the QTL matrix, target QTLs, and their individual breeding lines in their programs.

Incorporation of molecular marker-based selections into a conventional breeding program will require breeders to custom their molecular breeding schemes and tailor them directly into their specific breeding objectives. However, understanding molecular properties of the quality matrix requires tremendous investment and undoubtedly represents the current bottleneck as to why successful exploitation of available mapped QTLs into cultivar development remains limited at the present time. With the advancement in molecular techniques, such as high throughput SNP technology, developing a SNP chip to specifically guard the quality matrix will be possible. Once customized chips can be developed for individual breeding programs, any novel traits from the global source (different genetic backgrounds, inter or intra subspecies or wild-related ancestors from global populations) can potentially be incorporated into their breeding programs to add and/or improve specific quality or to boost yield without jeopardizing locally adapted standard qualities. Private companies have developed proprietary methodology that allows their breeders to combine their germplasm knowledge and breeding population objectives with molecular phenotypic trait association in order to develop genetic modeling for multiple marker-assisted selections and obtain rapid increase in the frequency of favorable alleles associated with target traits within the breeding population (Eathington et al., 2007).

7. References

- Abad-Grau, M.M.; Montes, R. & Sebastiani, P. (2006). Building chromosome-wide LD maps. *Bioinformatics* 22: 1933–1934, ISSN 1367-48
- Aggarwal, R.K.; Hendre, P.S., Varshney, R.K., Bhat, P.R., Krishnakumar, V. & Singh, L. (2007). Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. *Theoretical and Applied Genetics* 114(2):359-372, ISSN 0040-5752
- Allen-Brady, K.; Wong, J. & Camp, N.J. (2006). PedGenie: an analysis approach for genetic association testing in extended pedigrees and genealogies of arbitrary size. *BMC Bioinformatics* 7: 209, ISSN 1471-2105
- Alves-Freitas, D.M.T.; Kilian, A. & Grattapaglia, D. (2011). Development of DArT (Diversity Arrays Technology) for high-throughput genotyping of *Pinus taeda* and closely related species. *BMC Proceedings* 2011 5(Suppl 7):P22.
<http://www.biomedcentral.com/1753-6561/5/S7/P22>
- Ammiraju, J.S.S.; Luo, M., Goicoechea, J.L., Wang, W., Kudrna, D., Mueller, C., Talag, J., Kim, H., Sisneros, N.B., Blackmon, B., Fang, E., Tomkins, J.B., Brar, D., MacKill, D., McCouch, S., Kurata, N., Lambert, G., Galbraith, D.W., Arumuganathan, K., Rao, K., Walling, J.G., Gill, N., Yu, Y., SanMiguel, P., Soderlund, C., Jackson, S. & Wing, R.A. (2006). The *Oryza* bacterial artificial chromosome library resource: construction and analysis of deep-coverage large insert BAC libraries that represent the 10 genome types of genus *Oryza*. *Genome Research* 16(1):140-147, ISSN 1088-9051
- Ammiraju, J.S.S.; Luo, M., Sisneros, N., Angelova, A., Kudrna, D., Kim, H., Yu, Y., Goicoechea, J.L., Lorieux, M., Kurata, N., Brar, D., Ware, D., Jackson, S. & Wing, R.A. (2010). The *Oryza* BAC resource: a genus wide and genome scale tool for exploring rice genome evolution and leveraging useful genetic diversity from wild relatives. *Breeding Science* 60: 536–543, ISSN 1344-7610
- Andersen, J.R. & Lübberstedt, T. (2003). Functional markers in plants. *Trends in Plant Science* 8: 554–560, ISSN 1360-1385
- Baird, N.A.; Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A. & Johnson, E.A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3(10):e3376, ISSN 1932-6203
- Beavis, W.D. (1994). The power and deceit of QTL experiment: Lessons from comparative QTL studies. In. *Proc. Corn Sorghum Ind Res Conf*, pp. 250-266, ISBN, Chicago, IL., USA. Dec. 7-8, 1994
- Bernardo, R. (2008). Molecular markers and selection for complex traits in plants: Learning from the last 20 years. *Crop Science* 48:1649-1664, ISSN 0011-183X
- Bernardo, R. & Charcosset, A. (2006). Usefulness of gene information in marker-assisted recurrent selection: A simulation appraisal. *Crop Science* 46:614-621, ISSN 0011-183X
- Blanchard, J.L. (2004). Bioinformatics and systems biology, rapidly evolving tools for interpreting plant response to global change. *Field Crops Research* 90:117–131, ISSN 0378-4290
- Bordes, J.; Charmet, G., Dumas de Vaulx, R., Pollacsek, M., Beckert, M. & Gallais, A. (2006). Doubled haploid versus S1 family recurrent selection for testcross performance in a maize population, *Theoretical and Applied Genetics* 112: 1063–1072, ISSN 0040-5752

- Borevitz, J.O.; Maloof, J.N., Lutes, J., Dabi, T., Redfern, J.L., Trainer, G.T., Werner, J.D., Asami, T., Berry, C.C., Weigel, D. & Chory, J. (2002). Quantitative trait loci controlling light and hormone response in two accessions of *Arabidopsis thaliana*. *Genetics* 160:683–696, ISSN 0016-6731
- Campbell, B.T.; Baenziger, P.S., Gill, K.S., Eskridge, K.M., Budak, H., Erayman, M., Dweikat, I. & Yen, Y. (2003). Identification of QTLs and environmental interactions associated with agronomic traits on chromosome 3A of wheat. *Crop Science* 43:1493–1505, ISSN 0011-183X
- Chaib, J.; Lecomte, L., Buret, M. & Causse, M. (2006). Stability over genetic backgrounds, generations and years of quantitative trait locus (QTLs) for organoleptic quality in tomato. *Theoretical and Applied Genetics* 112:934–944, ISSN 0040-5752
- Choo, T.M.; Reinbergs, E. & Kasha, K.J. (1985). Use of haploids in breeding barley. *Plant Breeding Reviews* 3:219–252, ISSN 0730-2207
- CoGePedia. (2011). Sequenced plant genomes. (Verified: 11/21/2011)
http://genomevolution.org/wiki/index.php/Sequenced_plant_genomes
- Davey, J.W.; Hohenhole, P.A., Etter, P.D., Boone, J.O., Catchen, J.M., & Blaxter, M.L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* 12:499–510, ISSN 1471-0056
- Delseny, M.; Bin Han, B. & Hsing, Y.I. (2010). High throughput DNA sequencing: The new sequencing revolution. *Plant Science* 179:407–422, ISSN 0168-9452
- Eathington, S.R.; Crosbie, T.M., Edwards, M.D., Reiter, R. & Bull, J.K. (2007). Molecular markers in commercial breeding program. *Crop Science* 47(S3):S154–S163, ISSN 0011-183X
- Edwards, D.; Forster, J.W., Chagné, D. & Batley, J. (2007). What is SNPs? In: *Association Mapping in Plants*, Oraguzie, N.C., Rikkerink, E.H.A., Gardiner, S.E. and De Silva, H.N. (eds), pp. 41–52, Springer, ISBN 978-0-387-35844-4, Berlin, Germany
- Edwards, D. & Batley, J. (2010). Plant genome sequencing: applications for crop improvement. *Plant Biotechnol. J.* 7:2–9, ISSN 1467-7644
- Eujayl, I.; Sorrels, M.E., Baum, M., Wolters, P. & Powell, W. (2002). Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. *Theoretical and Applied Genetics* 104:399–407, ISSN 0040-5752
- Feuillet, C.; Leach, J.E., Rogers, J., Schnable, P.S. & Eversole, K. 2011. Crop genome sequencing: lessons and rationales. *Trends in Plant Science* 16(2):77–88, ISSN 1360-1385
- Forster, B.P. & Thomas, W.T.B. (2004). Doubled haploids in genetics and plant breeding. *Plant Breeding Reviews* 25:57–88, ISSN 0730-2207
- Frampton, A. (2008). Integration of commercial plant breeding and genomic technologies. United States Patent Application 20080034450
- Fridman, E.; Carrari, F., Liu, Y-S, Fernie, A.R. & Zamir, D. (2004). Zooming in on quantitative trait for the tomato yield using interspecific introgression. *Science* 305:1786–1789, ISSN 0036-8075
- Frisch, M.; Bohn, M. & Melchinger, A.E. (1999). Comparison of selection strategies for marker-assisted backcrossing of gene. *Crop Science* 39(5):1295–1301, ISSN 0011-183X
- Frisch, M. & Melchinger, A.E. (2005). Selection theory for marker-assisted backcrossing. *Genetics* 170: 909–917, ISSN 0016-6731

- Fukuoka, S.; Ebana, K., Yamamoto, T. & Yano, M. (2010). Integration of Genomics into Rice Breeding. *Rice* 3:131-137, ISSN 1939-8425
- Garcia, A.A.; Kido, E.A., Meza, A.N., Souza, H.M., Pinto, L.R., Pastina, M.M., Leite, C.S., Silva, J.A., Ulian, E.C., Figueira, A. & Souza, A.P. (2006). Development of an integrated genetic map of a sugarcane (*Saccharum* spp.) commercial cross, based on a maximum-likelihood approach for estimation of linkage and linkage phases. *Theoretical and Applied Genetics* 112:298-314, ISSN 0040-5752
- Gaunt, T.R.; Rodriguez, S., Zapata, C. & Day, I.N.M. (2006). MIDAS: software for analysis and visualization of interallelic disequilibrium between multiallelic markers. *BMC Bioinformatics* 7:227, ISSN 1471-2105
- Harlan, J.R. (1992). Crops and Man. American Society of Agronomy and Crop Science Society of America, Madison, WI, ISBN 0-89118-107-5
- Haupt, W.; Fischer T.C., Windler, S., Fransz, P. & Torres-Ruiz, R.A. (2001). The centromere1 (CEN1) region of *Arabidopsis thaliana*: Architecture and functional impact of chromatin. *Plant Journal* 27:285-296, ISSN 0960-7412
- Herzog, E. & Frisch, M. (2011). Selection strategies for marker-assisted backcrossing with high-throughput marker systems. *Theoretical and Applied Genetics* 123:251-260, ISSN 0040-5752, ISSN 1344-7610
- Hohenlohe, P.A.; Bassham, S., Etter, P.D., Stiffler, N., Johnson, E.A. & Cresko, W.A. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics* 6(2):e1000862, ISSN 1553-7390
- Hospital, F. (2001). Size of donor chromosome segments around introgressed loci and reduction of linkage drag in marker-assisted backcross programs. *Genetics* 2001; 158(3):1363-1379, ISSN 0016-6731
- Hospital, E. & Charcosset, A. (1997). Marker-assisted introgression of qualitative trait loci. *Genetics* 1997; 147(3):1469-1485, ISSN 0016-6731
- Howes, N.K.; Woods, S.M. & Townley-Smith, T.F. (1998). Simulations and practical problems of applying multiple marker-assisted selection and doubled haploid to wheat breeding programs. *Euphytica* 100:225-230, ISSN 0014-2336
- Hyten, D.L.; Cannon, S.B., Song, Q., Weeks, N., Fickus, E.W., Shoemaker, R.C., Specht, J.E., Farmer, A.D., May, G.D. & Cregan, P.B. (2010). High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. *BMC Genomics* 11:38 (<http://www.biomedcentral.com/1471-2164/11/38>)
- Jaccoud, D.; Peng, K., Feinstein, D. & Kilian, A. (2001). Diversity arrays: A solid state technology for sequence information independent genotyping. *Nucleic Acids Research* 29: e25, ISSN 1362-4962
- Jansen, R.C. & Beavis, W. (2001). MQM Mapping using haplotyped putative QTL-alleles; a simple approach for mapping QTL's in plant breeding populations. Patent EP 1265476
- Jenkins, S. & Gibson, N. (2002). High-throughput SNP genotyping. *Comparative and Functional Genomics* 3, 57-66, ISSN 1532-6268
- Jeung, J.U.; Heu, S.G., Shin, M.S., Vera Cruz, C.M. & Jena, K.K. (2006). Dynamics of *Xanthomonas oryzae* pv populations in Korea and their relationship to known bacterial blight resistant genes. *Phytopathology* 96(8):867-875, ISSN 0031-949X
- Jourjon, M.F., Jasson, S., Marcel, J., Ngom, B. & Mangin, B. (2005). MCQTL: multi-allelic QTL mapping in multi-cross design. *Bioinformatics* 21:128-130, ISSN 1367-48

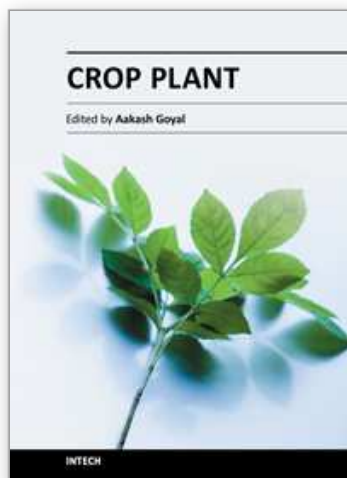
- Khatri, P. & Draghici, S. (2005). Ontological analysis of gene expression data: current tools, limitations and open problems. *Bioinformatics* 21:3587–3595, ISSN 1367-48
- Khatri, P.; Draghici, S., Ostermeier, G.C. & Krawetz, S.A. (2002) Profiling gene expression using OntoExpress. *Genomics* 79:266–270, ISSN 0888-7543
- Koebner, R. (2003). MAS in cereals: Green for maize, amber for rice, still red for wheat and barley. In: *Marker assisted selection: A fast tract to genetic gain in plant and animal breeding?* 17-18 Oct. 2003; pp. 12-17, Turin, Italy. FAO, Rome. (www.fao.org/biotech/docs/Koebner.pdf)
- Lander, E.S.; Green, P., Abrahamson, J., Barlow, A., Daly, M.J., Lincoln, S.E. & Newburg, L. (1987). MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1:174–181, ISSN 0888-7543
- Levinson, G. & Gutman, G.A. (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution* 4:203–221, ISSN 0737-4038
- Lewis, Z.A.; Shiver, A.L., Stiffler, N., Miller, M.R., Johnson, E.A. & Selker, E.U. (2007). High density detection of restriction site associated DNA (RAD) markers for rapid mapping of mutated loci in *Neurospora*. *Genetics* 177(2):1163–1171, ISSN 0016-6731
- Li, F.; Kitashiba, H., Inaba, K. & Nishio, T. (2009). A *Brassica rapa* linkage map of EST-based SNP markers for identification of candidate genes controlling flowering time and leaf morphological traits, *DNA Research* 16:311–23, ISSN 1756-1663
- Liu, J.S.; Sabatti, C., Teng, J., Keats, B.J.B. & Risch, K. (2001). Bayesian analysis of haplotypes for linkage disequilibrium mapping. *Genome Research* 11:1716–1724, ISSN 1088-9051
- Lu, X.; Niu, T. & Liu, J.S. (2003). Haplotype information and linkage disequilibrium mapping for single nucleotide polymorphisms. *Genome Research* 13:2112–2117, ISSN 1088-9051
- Lübberstedt, T.; Melchenger, A.E., Fähr, S., Klein, D., Dally, A. & Westhoff, P. (1998). QTL mapping in test crosses of flint lines of maize: III. Comparison across populations for forage traits. *Crop Science* 38:1278–1289
- Manly, K.F. (1993). A Macintosh program for storage and analysis of experimental genetic mapping data. *Mammalian Genome* 4:303–313, ISSN 0938-8990
- Martinez, V.; Thorgaard, G., Robison, B. & Sillanpää, M.J. (2005) An application of Bayesian QTL mapping to early development in double haploid lines of rainbow trout including environmental effects. *Genetical Research* 86:209–221, ISSN 0016-6723
- Matsumoto, T.; Wu J.Z., Kanamori, H., et al. (2005). The map-based sequence of the rice genome. *Nature* 436:793–800, ISSN 0028-0836
- Mayes, S.; Parsley, K., Sylvester-Bradley, R., May, S. & Foulkes, J. (2005). Integrating genetic information into plant breeding programs: how will we produce varieties from molecular variation, using bioinformatics? *Annals of Applied Biology* 146:223–237, ISSN 0003-4746
- McCouch S.R.; Teytelman, L., Xu, Y., Lobos, K.B, Clare, K., Walton, M., Fu, B., Maghirang, R., Li, Z., Xing, Y., Zhang, Q., Kono, I., Yano, M., Fjellstrom, R., DeClerck, G.G., Schneider, D., Cartinhour, S., Ware, D. & Stein, L. (2002). Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.), *DNA Research* 9(6):199–207, ISSN 1340-2838
- McCouch, S.R.; Zhao, K., Wright, M., Tung, C.W, Ebana, K., Thomson, M., Reynolds, A., Wang, D., DeClerck, G., Ali, M.L., McClung, A., Eizenga, G. & Bustamante, C.

- (2010). Development of genome-wide SNP assays for rice, *Breeding Science* 60: 524–53, ISSN 1344-7610
- McNally, K.L.; Bruskiewich, R., Mackill, D., Buell, C.R., Leach, J.E. & Leung, H. (2006). Sequencing multiple and diverse rice varieties. Connecting whole genome variation with phenotypes, *Plant Physiol.* 141:26–33, ISSN 0032-0889
- McNally, K.L.; Childs, K.L., Bohnert, R., Davidson, R.M., Zhao, K., Ulat, V.J., Zeller, G., Clark, R.M., Hoen, D.R., Bureau, T.E., Stokowski, R., Ballinger, D.G., Frazer, K.A., Cox, D.R., Padhukasahasram, B., Bustamante, C.D., Weigel, D., Mackill, D.J., Bruskiewich, R.M., Ratsch, G., Buell, C.R., Leung, H. & Leach, J.L. (2009) Genome-wide SNP variation reveals relationships among landraces and modern varieties of rice. *Proc. Natl. Acad. Sci. USA* 106:12273–12278, ISSN-0027-8424
- Miller, M.R.; Atwood, T.S., Eames, B.F., Eberhart, J.K., Yan, Y.L., Postlethwait, J.H. & Johnson, E.A. (2007a). RAD marker microarrays enable rapid mapping of zebrafish mutations. *Genome Biololgy.* 8(6):R105, ISSN 1465-6906
- Miller, M.R.; Dunham, J.P., Amores, A., Cresko, W.A. & Johnson, E.A. (2007b). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research* 17(2):240-248, ISSN 1088-9051
- Mochida, K. & Shinozak, K. (2010). Genomics and Bioinformatics Resources for Crop Improvement. *Plant Cell Physiology* 51(4):497-523, ISSN 0032-0781
- Monforte, A.J. & Tanksley, S.D. (2000). Fine mapping of a quantitative trait locus (QTL) from *Lycopersicon hirsutum* chromosome 1 affecting fruit characteristics and agronomic traits: Breaking linkage among QTLs affecting different traits and dissection of heterosis for yield. *Theoretical and Applied Genetics* 100:471–479, ISSN 0040-5752
- Nas, T.M.S.; Sanchez, D.L., Diaz, G.Q., Mendioro, M.S. & Virmani, S.S. (2005). Pyramiding of thermosensitive genetic male sterility (TGMS) genes and identification of candidate tms5 gene in rice. *Euphytica* 145(1-2):67-75, ISSN 0014-2336
- Oliveira, K.M.; Pinto, L.R., Marconi, T.G., Mollinari, M., Ulian, E.C., Chabregas, S.M., Falco, M.C., Burnquist, W., Garcia, A.A.F. & Souza, A.P. (2009). Characterization of ne polymorphic functional markers for sugarcane. *Genome* 52:191-209, ISSN 0831-2796
- Paux, E.; Sourdille, P., Salse, J., Saintenac, C., Choulet, F., Leroy, P., Korol, A., Michalak, M., Kianian, S., Spielmeier, W., Lagudah, E., Somers, D., Kilian, A., Alaux, M., Vautrin, S., Berges, H., Eversole, K., Appels, R., Safar, J., Simkova, H., Dolezel, J., Bernard, M. & Feuillet, C. (2008). A physical map of the 1-gigabase bread wheat chromosome 3B. *Science* 322:101-104, ISSN 0036-8075
- Primmer, C.R.; Ellengren, H., Saino, N. & Moller, A.P. (1996). Directional evolution in germline microsatellite mutations. *Nature Genetics* 13:391–393, ISSN 1061-4036
- Pritchard, J.K.; Stephens, M., Rosenberg, N.A. & Donnelly, P. (2000). Association mapping in structured populations. *American Journal of Human Genetics* 67:170–181, ISSN 0002-9297
- Purcell, S.; Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., de Bakker, P.I.W.; Daly, M.J. & Sham, P.C. (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics* 81:559–575, ISSN 0002-9297
- Ramchiary, N.; Nguyen, V.D., Li, X., Hong, C.P., Dhandapani, V., Choi, S.R., Yu, G., Piao, Z.Y. & Lim, Y.P. (2011). Genic microsatellite markers in *Brassica rapa*: Development, characterization, mapping, and their utility in other cultivated and

- wild Brassica relatives, *DNA Research* pp. 1–16, doi:10.1093/dnares/dsr017, ISSN 1756-1663
- Seaton, G.; Haley, C.S., Knott, S.A., Kearsey, M. & Visscher, P.M. (2002) QTL Express: mapping quantitative trait loci simple and complex pedigrees. *Bioinformatics* 18:339–340, ISSN 1367-4803
- Semagn, K., Bjørnstad, A. & Xu Y. (2010). The genetic dissection of quantitative traits in crops. *Electronic Journal of Biotechnology* 13:5, <http://dx.doi.org/10.2225/vol13-issue5-fulltext-14>, ISSN 0717-3458
- Shen, Y.-J.; Jiang, H., Jin, J.-P., Zhang, Z.-B., Xi, B., He, Y.-Y., Wang, G., Wang, C., Qian, L., Li, X., Yu, Q.-B., Liu, H.-J., Chen, D.-H., Gao, J.-H., Huang, H., Shi, T.-L. & Yang, Z.-N. (2004). Development of genome-wide DNA polymorphism database for map-based cloning of rice genes. *Plant Physiology* 135:1198–1205, ISSN 0032-0889
- Shirasawa, K.; Oyama, M., Hirakawa, H., Sato, S., Tabata, S., Fujioka, T., Kimizuka-Takagi, C., Sasamoto, S., Watanabe, A., Kato, M., Kishida, Y., Kohara, M., Takahashi, C., Tsuruoka, H., Wada, T., Sakai, T. & Isobe, S. (2011). An EST-SSR linkage map of *Raphanus sativus* and comparative genomics of the Brassicaceae. *DNA Research* 18:221–232, ISSN 1756-1663
- Singer, T.; Fan, Y., Chang, H.S., Zhu, T., Hazen, S.P., & Briggs, S.P. (2006). A high-resolution map of *Arabidopsis* recombinant inbred lines by whole-genome exon array hybridization. *PLoS Genet* 2(9): e144. DOI: 10.1371/journal.pgen.0020144, ISSN 1932-6203
- Skol, A.D.; Scott, L.J., Abecasis, G.R. & Boehnke, M. (2006). Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nature Genetics* 38:209–213, ISSN 1061-4036
- Smith, O.S.; Cooper, M., Tingey, S.V., Rafalski, A.J., Luedtke, R. & Niebur, W.S. (2005). Plant breeding method, WIPO Patent Application WO05000006
- Sonah, H.; Deshmukh, R.K., Sharma, A., Singh, V.P., Gupta, D.K., Gacche, R.N., Rana, J.C., Singh, N.K. & Sharma, T.R. (2011). Genome-wide distribution and organization of microsatellites in plants: an insight into marker development in *Brachypodium*. *PLoS ONE* 6(6):e21298. doi:10.1371/journal.pone.0021298, ISSN 1932-6203
- Tan, L.; Liu, F., Xue, W., Wang, G., Ye, S., Zhu, Z., Fu, Y., Wang, X. & Sun, C. (2007). Development of *Oryza rufipogon* and *O. sativa* introgression lines and assessment for yield related quantitative trait loci. *Journal of Integrative Plant Biology*, 49: 871–884
- Tautz, D. & Renz, M. (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Research* 12:4127–4138
- Thiel, T.; Michalek, W., Varshney, R.K. & Graner, A. (2003). Exploiting EST data bases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics* 106:411–422, ISSN 0040-5752
- Ukai, Y.; Osawa, R., Saito, A. & Hayashi, T. (1995). MAPL: a package of computer programs for construction of DNA polymorphism linkage maps and analysis of QTL (in Japanese). *Breeding Science* 45:139–142, ISSN 1344-7610
- Utz, H.F. & Melchinger, A.E. (1996). PLABQTL: a program for composite interval mapping of QTL. *Journal of Agricultural Genomics* 2(1), Available at: www.cabi-publishing.org/jag/papers96/paper196/indexp196.html

- Van Ooijen, J.W. & Voorrips, R.E. (2001). JoinMap® version 3.0: software for the calculation of genetic linkage maps. Wageningen: Plant Research International
- Varshney, R.K.; Graner, A. & Sorrells, M.E. (2005). Genic microsatellite markers in plants: features and applications. *Trends in Biotechnology* 23:48–55, ISSN 0167-7799
- Varshney, R.K.; Nayak, S.N., May, G.D. & Jackson, S.A. (2009). Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends in Biotechnology* 27: 522 – 530, ISSN 0167-7799
- Vigouroux, Y.; Mitchell, S., Matsuoka, Y., Hamblin, M., Kresovich, S., Smith, J.S.C., Jaqueth, J., Smith, O.S. & Doebley, J. (2005). An analysis of genetic diversity across the maize genome using microsatellites. *Genetics* 169:1617–1630, ISSN 0016-6731
- Vision, T.J.; Brown, D.G., Shmoys, D.B., Durrett, R.T. & Tanksley, S.D. (2000). Selective mapping: a strategy for optimizing the construction of high-density linkage maps. *Genetics* 155:407–420, ISSN 0016-6731
- Wang, J.; Chapman, S.C., Bonnet, D.G., Rebetzke, G.J. & Crouch, J. (2007a). Application of population genetic theory and simulation models to efficiently pyramid multiple genes via marker-assisted selection. *Crop Science* 47:582–588, ISSN 0011-183X
- Wang, Z.; Jia, Y., Rutger, J.N. & Xia, Y. (2007b). Rapid survey for presence of a blast resistance gene Pi-ta in rice cultivars using the dominant DNA markers derived from portions of the Pi-ta gene. *Plant Breeding* 126(1): 36–42, ISSN 0179-9541
- Wang, Z.; Li, J., Luo, Z., Huang, L., Chen, X., Fang, B., Li, Y., Chen, J. & Zhang, X. (2011). Characterization and development of EST-derived SSR markers in cultivated sweet potato (*Ipomoea batatas*). *BMC Plant Biology* 2011, 11:139, ISSN 1471-2229
- Ware, D.H.; Jaiswal, P., Ni, J., Yap, I.V., Pan, X., Clark, K.Y., Teytelman, L., Schmidt, S.C., Zhao, W., Chang, K., Cartinhour, S., Stein, L.D. & McCouch, S.R. (2002) Gramene, a tool for grass genomics. *Plant Physiology* 130:1606–1613, ISSN: 0032-0889
- Wei, X.M.; Jackson, P.A., Hermann, S., Kilian, A., Heller-Uszynska, K., & Deomano, E. (2010). Simultaneously accounting for population structure, genotype by environment interaction, and spatial variation in marker-trait associations in sugarcane. *Genome* 2010, 53(11):973–981, ISSN 0831-2796
- Wei, B.; Jing, R., Wang, C., Chen, J., Mao, X., Chang, X. & Jia, J. (2009). Dreb1 genes in wheat (*Triticum aestivum* L.): development of functional markers and gene mapping based on SNPs. *Molecular Breeding* 23: 13–22, ISSN 1380-3743
- Wenzl, P.; Carling, J., Kudrna, D., Jaccoud, D., Huttner, E., Kleinjohs, A. & Kilian, A. (2004). Diversity Arrays Technology (DArT) for whole-genome profiling of barley, *PNAS* 101 (26):9915–9920, ISSN-0027-8424
- Wing, R.A.; Ammiraju, J.S.S., Luo, M., Kim, H., Yu, Y., Kudrna, D., Goicoechea, J.L., Wang, W., Nelson, W., Rao, K., Brar, D., Mackill, D.J., Han, B., Soderlund, C., Stein, L., SanMiguel, P. & Jackson, S. (2005). The *Oryza* map alignment project: The golden path to unlocking the genetic potential of wild rice species. *Plant Molecular Biology* 59(1):53–62, ISSN 0167-4412
- Wing, R.; Kim, H., Foicoechea, J., Yu, Y., Kudrna, D., Zuccolo, A., Ammiraju, J., Luo, M., Nelson, W. & Ma, J. (2007). The *Oryza* map alignment project (OMAP): a new resource for comparative genome studies within *Oryza*. In: Upadhyaya, N.M. (ed.) *Rice Functional Genomics*, Springer, New York, pp. 395–409. ISBN 978-0-387-48903-2

- Xia, L.; Peng, K., Yang, S., Wenzl, P., de Vicente, C., Fregene, M., Kilian, A. (2005). DArT for high-throughput genotyping of cassava (*Manihot esculenta*) and its wild relatives. *Theoretical and Applied Genetics* 110:1092–1098, 411–422, ISSN 0040-5752
- Xu, S. (2003a). Theoretical basis of the Beavis effect. *Genetics* 165:2259–2268, ISSN 0016-6731
- Xu, Y. (2003b). Developing marker-assisted selection strategies for breeding hybrid rice. *Plant Breeding Reviews* 23:73–174, ISSN 0730-2207
- Xu, Y., McCouch, S.R. & Zhang, Q. (2005). How can we use genomics to improve cereals with rice as a reference genome? *Plant Molecular Biology* 59:7–26, ISSN 0167-4412
- Xu, Y. & Crouch, J.H. (2008). Marker-assisted selection in plant breeding: From publication to practice. *Crop Science* 48:391–407, ISSN 0011-183X
- Yan, J., Yang, X., Shah, T., Sánchez-Villeda, H., Li, J., Warburton, M., Zhou, Y., Crouch, J.H. & Xu, Y. (2010). High-throughput SNP genotyping with the GoldenGate assay in maize. *Molecular Breeding* 25:441–451, ISSN 1380-3743
- Yang, J., Hu, C., Hu, H., Yu, R., Xia, Z., Ye, X. & Zhu, J. (2008). QTLNetwork: mapping and visualizing genetic architecture of complex traits in experimental populations. *Bioinformatics* 24:721–723, ISSN 1367-48
- Yang, S.Y.; Saxena, R.K., Kulwal, P.L., Ash, G.J., Dubey, A., Harpe, J.D.I., Upadhyaya, H.D., Gothwal, R., Kilian, A. & Varshney, R.K. (2011). The first genetic map of pigeonpea based on Diversity Arrays Technology (DArT) markers. *Journal of Genetics* 90(1):103–109, ISSN 0022-1333
- Yoshimura, A.; Lei, J.X., Matsumoto, T., Tsunematsu, H., Yoshimura, S., Iwata, N., Baraoidan, M.R., Mew, T.W. & Nelson, R.J. (1996). Analysis and pyramiding of bacterial blight resistance genes in rice by using DNA markers. 577–581. In: *Rice genetics III. Proceedings of the third international rice genetics symposium. International Rice Research Institute, G.S. Khush (ed.)*, pp. 577–581, ISBN 971-22-0087-6, Manila, Philippines
- Yu, H.; Xie, W., Wang, J., Xing, Y., Xu, C., Li, X., Xiao, J. & Zhang, Q. (2011). Gains in QTL detection using an ultra-high density SNP map based on population sequencing relative to traditional RFLP/SSR markers. *PLoS ONE* 6(3):e17595. doi:10.1371/journal.pone.0017595, ISSN1932-6203
- Yu, J.K.; La Rota, M., Kantety, R.V. & Sorrells, M.E. (2004). EST-derived SSR markers for comparative mapping in wheat and rice. *Molecular Genetics and Genomics* 271:742–751, ISSN 1617-461
- Yu, J.; Wang, J., Lin, W., Li, S., Li, H., et al. (2005). The genomes of *Oryza sativa*: A history of duplications. *PLoS Biol* 3(2): e38. DOI: 10.1371/journal.pbio.0030038, ISSN 1932-6203
- Zhang, Z.; Bradbury, P.J., Kroon, D.E., Casstevens, T.M. & Buckler, E.S. (2006). *TASSEL 2.0: a software package for association and diversity analyses in plants and animals*. Poster presented at Plant and Animal Genomes XIV Conference, 14–18 January 2006, San Diego, California



Crop Plant

Edited by Dr Aakash Goyal

ISBN 978-953-51-0527-5

Hard cover, 240 pages

Publisher InTech

Published online 20, April, 2012

Published in print edition April, 2012

This book provides us a thorough overview of Crop Plant with current advance in research. Divided into two section based on the chapters contents. Chapter 1 provides information about markers and next generation sequencing technology and its use. Chapter 2 is about how we can use Silicon for Drought tolerance. Chapter 3 is to deal with the major problem of rising CO₂ and O₃ causing environmental pollution. Chapter 4 covers the phenomena of RNAi and its use, application in crop science. Chapter 5 is a review for boron deficiency in soils and how to deal with it for better crops. Chapter 6-10 provide some information regarding recent works going on in crop science.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Herry S. Utomo, Ida Wenefrida and Steve D. Linscombe (2012). Progression of DNA Marker and the Next Generation of Crop Development, Crop Plant, Dr Aakash Goyal (Ed.), ISBN: 978-953-51-0527-5, InTech, Available from: <http://www.intechopen.com/books/crop-plant/progression-of-dna-marker-and-the-next-generation-of-crop-development>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IntechOpen

IntechOpen