

Chapter 2

Quantitative Trait Loci Mapping in Plants: Concepts and Approaches

Deepmala Sehgal, Richa Singh and Vijay Rani Rajpal

Abstract The narrow genetic base of modern crop cultivars is a serious obstacle to sustain and improve crop productivity due to rapidly occurring vulnerability of genetically uniform cultivars to potentially new biotic and abiotic stresses. Plant germplasm resources, originated from a number of historical genetic events as a response to environmental stresses and selection, are the important reservoirs of natural genetic variations that can be exploited to increase the genetic base of the cultivars. However, many agriculturally important traits such as productivity and quality, tolerance to environmental stresses, and some of forms of disease resistance are quantitative (also called polygenic, continuous, multifactorial, or complex traits) in nature. The genetic variation of a quantitative trait is controlled by the collective effects of numerous genes, known as quantitative trait loci (QTLs). Identification of QTLs of agronomic importance and its utilization in a crop improvement requires mapping of these QTLs in the genome of crop species using molecular markers. This review will focus on the basic concepts and a brief description of existing methodologies for QTL mapping and their merits and demerits including traditional biparental mapping and the advanced linkage disequilibrium (LD)-based association mapping. Examples of some of the recent studies on association mapping in various crop species are provided to demonstrate the merits of high-resolution association mapping approach over traditional mapping methods. This review thus will provide non-expert readers of crop breeding community an opportunity to develop a basic understanding of dissecting and exploiting natural variations for crop improvement.

D. Sehgal (✉)

International Maize and Wheat Improvement Center (CIMMYT), Km.45,
Carretera, Mex-Veracruz El Batán, Texcoco CP 56130, Edo. de México, Mexico
e-mail: d.sehgal@cgiar.org

R. Singh

Department of Molecular Biology and Biotechnology, Chaudhary Charan Singh Haryana
Agricultural University (CCSHAU), Hisar, Haryana 125004, India
e-mail: richabiotech77@gmail.com

V.R. Rajpal

Department of Botany, HansRaj College, University of Delhi, Delhi 11007, India
e-mail: vijayrani2@gmail.com

Keywords Linkage map · Quantitative trait loci · Linkage disequilibrium · Association mapping

2.1 Introduction

As most of the modern genotypes cultivated today have descended from a relatively small number of landraces, the genes controlling important traits have reduced diversity compared to the gene pool of landraces and wild relatives. In view of such threats of constant genetic erosion, gene banks have been established and maintained in order to mainly preserve wild plant accessions as well as landraces. Gene bank collections are the important reservoirs of natural genetic variations originating from a number of historical genetic events as a response to environmental stresses (Hoisington et al. 1999). Unlocking biodiversity held in gene banks and mobilizing useful variation to breeding programs are required for the genetic improvement of crops and to meet the overarching goal of diversification of the adapted gene pools. However, many agriculturally important variations such as productivity and quality, tolerance to environmental stresses, and some of forms of disease resistance are controlled by polygenes which complicate the breeding process since phenotypic performances only partially reflects the genetic values of individuals. These complex traits are referred to as quantitative traits (also called as polygenic or multifactorial traits) and the regions within genomes that contain genes associated with a particular quantitative trait are known as quantitative trait loci (QTLs).

It is challenging to identify QTLs based on only traditional phenotypic evaluation. Identification of QTLs of agronomic importance and its utilization in a crop improvement requires mapping of these QTLs in a genome of crop species using molecular markers. Identification of QTLs with genetically linked DNA-markers is useful for incorporating genes into improved cultivars via marker-assisted selection (MAS), map-based cloning of the tagged genes, and for a better understanding of the genetics of complex traits (Asins 2002). Linkage analysis and association mapping are the two most commonly used methods for QTL mapping. This review provides an overview of the two QTL mapping methods with special focus on mapping population types, linkage map construction, and marker–trait association analysis using different statistical methods and software programs. The contents of the review will provide non-expert readers of crop breeding community an opportunity to develop a basic understanding of dissecting and exploiting natural variations for crop improvement.

2.2 Principle of QTL Analysis

Identifying a QTL or a gene within a plant genome is like finding the proverbial needle in a haystack. However, QTL analysis can be used to divide the haystack in manageable piles and systematically search them. In simple terms, QTL analysis is based on the principle of detecting an association between phenotype and the genotype of markers. Markers are used to partition the mapping population into different genotypic groups based on the presence or absence of a particular marker locus and to determine whether significant differences exist between groups with respect to the trait being measured (Tanksley 1993; Young 1996). A significant difference between phenotypic means of the groups (either 2 or 3), depending on the marker system and type of population, indicates that the marker locus being used to partition the mapping population is linked to a QTL controlling the trait.

2.3 Traditional QTL Mapping (Linkage Mapping)

The general steps involved in a traditional QTL mapping experiment are as follows: (1) select two parental strains that have differences between them in the alleles that affect variation in a trait. The parents need not be different in the mean phenotypic value of the trait as different allelic combinations can yield the same phenotypic mean; (2) develop an appropriate mapping population by crossing the selected parents; (3) phenotype the mapping population for the trait(s) of interest (morphological characters, agronomic traits, disease and pest scores, drought resistance, etc.) under greenhouse, screen-house, and/or field conditions; (4) generate the molecular data on the population with adequate number of uniformly spaced polymorphic markers; (5) construct a genetic map; and (6) identify molecular markers linked to the trait(s) of interest using statistical programs.

2.3.1 Mapping Populations Used in QTL Mapping Experiments

Various types of mapping population may be produced from the heterozygous F_1 hybrids:

1. Double haploid lines (DHLs): plants are regenerated from pollen (which is haploid) of the F_1 plants and treated to restore diploid condition in which every locus is homozygous. Since the pollen population has been generated by meiosis, the DHLs represent a direct sample of the segregating gametes.
2. Backcross (BC) population: the F_1 plants are backcrossed to one of the parents.
3. F_2 population: F_1 plants are selfed.

4. $F_{2:3}/F_{2:4}$ lines: $F_{3/4}$ plants tracing back to the same F_2 plant, also called F_2 families.
5. Recombinant inbred lines (RILs): inbred generation derived by selfing individual F_2 plants and further single seed descent. A population of RILs represents an 'immortal' or permanent mapping population.

Each of the above mapping populations possesses advantages and disadvantages. Hence, the choice of the type of mapping population depends on many factors such as the plant species, type of marker system used, and the trait to be mapped later on. F_2 populations, derived from F_1 hybrids, and BC populations, derived by crossing the F_1 hybrid to one of the parents, are the simplest types of mapping populations developed for self-pollinating species. Their main advantages are that they are easy to construct and require only a short time to produce. However, such populations are not fixable due to their inherent heterozygous genetic constitution. This restricts their wide utility in QTL analysis. The length of time needed for producing RIL population is the major disadvantage, because usually six to eight generations are required. Development of a DH population takes much less time than RIL; however, the production of DH populations is only possible in species that are amenable to tissue culture (e.g., cereal species such as rice, barley, and wheat). The major advantages of RIL and DH populations are that they produce homozygous or 'true-breeding' lines that can be multiplied and reproduced without genetic change occurring. This allows for the conduct of replicated trials across different locations and years. Furthermore, seed from individual RIL or DH lines may be transferred between different laboratories for further linkage analysis and the addition of markers to existing maps. Information provided by co-dominant markers is best exploited by an F_2 population, while information obtained by dominant marker systems can be maximized by using RILs or DHLs. Double haploids, F_2 or F_3 families, or RILs are advantageous if the trait to be mapped cannot be accurately measured on a single-plant basis but must be assessed in replicated field experiments.

2.3.2 Construction of Genetic/Linkage Maps

A linkage map may be thought of as a 'road map' of the chromosomes derived from two different parents. Linkage maps indicate the position and relative genetic distances between markers along chromosomes. Construction of a linkage map, using genotyping data generated on any of the above-mentioned mapping populations, is an important step before initiating any QTL analysis. In a segregating mapping population, there is a mixture of parental and recombinant genotypes. The frequency of recombinant genotypes is used to calculate recombination fractions, which is then used to infer the genetic distance between markers. By analyzing the segregation of markers, the relative order and distances between markers can be determined; the lower the frequency of recombination between two markers, the

closer they are situated on a chromosome (conversely, the higher the frequency of recombination between two markers, the further away they are situated on a chromosome). Two commonly used mapping functions that convert recombination frequency into centimorgan (cM) distance are the Kosambi mapping function, which assumes that recombination events influence the occurrence of adjacent recombination events, and the Haldane mapping function, which assumes no interference between crossover events. Linkage between markers is usually calculated with an odds ratio (i.e., the ratio of linkage versus no linkage). This ratio is more conveniently expressed as the logarithm of the ratio and is called a logarithm of odds (LOD) value or LOD score (Risch 1992). LOD values of >3 are typically used to construct linkage maps. LOD values may be lowered in order to detect linkage over a greater distance or to place additional markers within maps constructed at higher LOD values (Collard et al. 2005). Linked markers are grouped together into linkage groups, which represent chromosomal segments or entire chromosomes.

It is important to note that distance on a linkage map is not directly related to the physical distance of DNA between genetic markers, but depends on the genome size of the plant species (Paterson 1996). Furthermore, the relationship between genetic and physical distance varies along a chromosome (Tanksley et al. 1992; Young 1994; Künzel et al. 2000). For example, there are recombination ‘hot spots’ and ‘cold spots,’ which are chromosomal regions in which recombination occurs more frequently or less frequently, respectively (Faris et al. 2000; Ma et al. 2001; Yao et al. 2002). Commonly used software programs for constructing linkage maps include Mapmaker/EXP (Lander et al. 1987; Lincoln et al. 1993), MapManager QTX (Manly et al. 2001), and THREaD Mapper Studio (Cheema et al. 2010) which are freely available from the Internet. JoinMap is another commonly used program for constructing linkage maps (Stam 1993).

2.3.3 *Detection of QTLs*

Four widely used methods for detecting QTLs are single-marker analysis, interval mapping by maximum likelihood, interval mapping by regression, and composite interval mapping.

2.3.3.1 **Single-Marker Analysis (Point Analysis)**

The traditional method to detect a QTL in the vicinity of a marker is studying single-genetic markers one at a time. The phenotypic means for progeny of each marker class are compared (e.g., means of the marker classes AA, Aa, aa). The difference between two means provides an estimate of the phenotypic effect of substituting an A allele by an a allele at the QTL. To test whether the inferred phenotypic effect is significantly different from zero, a simple statistical test, such as

t-test or *F*-test, is used. A significant value indicates that a QTL is located in the vicinity of the marker. Single-point analysis does not require a complete molecular linkage map. The farther a QTL is from the marker, the less likely it is to be detected statistically due to crossover events between the marker and the gene.

2.3.3.2 Interval Mapping by Maximum Likelihood

QTL interval mapping is the most common method of QTL analysis. The principle behind interval mapping is to test a model for the presence of a QTL at many positions between two mapped marker loci. This model is a fit, and its goodness is tested using the method of maximum likelihood. For example, if it is assumed that a QTL is located between two markers, the 2-loci marker genotypes (i.e., AABB, AAbb, aaBB, aabb for DH progeny) each contain mixtures of QTL genotypes. Maximum likelihood involves searching for QTL parameters that give the best approximation for quantitative trait distributions that are observed for each marker class. Models are evaluated by computing the likelihood of the observed distributions with and without fitting a QTL effect. The map position of a QTL is determined as the maximum likelihood from the distribution of likelihood values (LOD scores: ratio of likelihood that the effect occurs by linkage to likelihood that the effect occurs by chance), calculated for each locus.

2.3.3.3 Interval Mapping by Regression

Interval mapping by regression (Haley and Knott 1992) was developed primarily as a simplification of the maximum likelihood method. It is essentially the same as the method of basic QTL analysis (regression on coded marker genotypes) except that phenotypes are regressed on QTL genotypes. Since the QTL genotypes are unknown, they are replaced by probabilities estimated from the nearest flanking markers.

2.3.3.4 Composite Interval Mapping

One of the factors that weakens interval mapping is fitting the model for a QTL at only one location. There are two problems with this approach: (a) the effects of additional QTL will contribute to sampling variance and (b) if two QTLs are linked, their combined effects will cause biased estimates. The method of composite interval mapping (CIM) was proposed as solution (Jansen and Stam 1994; Utz and Melchinger 1996; Zeng 1994). CIM will perform the analysis in the usual way, except that the variance from other QTLs is accounted for by including partial regression coefficients from markers ('cofactors') in other regions of the genome. CIM gives more power and precision than simple interval mapping (SIM) because the effects of other QTLs are not present as residual variance. CIM can remove the bias that can be caused by QTLs that are linked to the position being tested.

2.4 QTL Mapping Softwares

There are over 100 genetic analysis software packages (linkage analysis and QTL mapping). Here, we list some features of the most commonly used software packages.

2.4.1 *MapMaker/QTL* (<ftp://genome.wi.mit.edu/pub/mapmaker3/>)

A user-friendly, freely distributed software program runs on almost all platforms. It analyzes F_2 or backcross data using standard interval mapping.

2.4.2 *MQTL*

MQTL is a computer program for CIM in multiple environments. It can also perform SIM. Currently, MQTL is restricted to the analysis of data from homozygous progeny (double haploids, or RILs). Progeny types with more than two marker classes (e.g., F_2) are not handled.

2.4.3 *PLABQTL* (<http://www.uni-hohenheim.de/~ipspwww/soft.html>)

PLABQTL is a freely distributed computer program for CIM and SIM of QTL. Its main purpose is to localize and characterize QTL in mapping populations derived from a biparental cross by selfing or production of double haploids. Currently, this program is the easiest software for composite interval mapping.

2.4.4 *QTL Cartographer* (<http://statgen.mcsu.edu/qtlcart/cartographer.html>)

QTL Cartographer is a QTL software written for either UNIX, Macintosh, or Windows. It performs single-marker regression, interval mapping, and composite interval mapping. It permits analysis from F_2 or backcross populations. It displays map positions of QTLs using the GNU PLOT software.

2.4.5 MapQTL (<http://www.cpro.dlo.nl/cbw/>)

MapQTL is a licensed software program. It performs Kruskal–Wallis test (single-marker analysis), CIM, and multiple interval mapping on almost all kinds of mapping populations.

2.4.6 Qgene

Qgene is a QTL mapping and marker-aided breeding package written for Macintosh. It has a user-friendly graphical interface and produces graphical outputs. QTL mapping is conducted by either single-marker regression or interval regression.

2.4.7 SAS

SAS is a general statistical analysis software. It can detect QTL by identifying associations between marker genotype and quantitative trait phenotype by single-marker analysis approach such as ANOVA, *t*-test, GLM, or REG.

Table 2.1 summarizes QTL mapping studies in three most important staple crops, viz. wheat, maize, and rice, for various traits using different marker systems, analysis procedure, and software programs.

2.5 Why Association Mapping?

The precision of QTL mapping largely depends on the genetic variation (or genetic background) covered by a mapping population, the size of a mapping population, and number of marker loci used. Once QTLs affecting a trait of interest are accurately tagged using above-outlined approach, marker tags are the most effective tools in a crop improvement that allows the mobilization of the genes of interest from donor lines to the breeding material through MAS. Although traditional QTL mapping will continue being an important tool in gene tagging of crops, it is a ‘now classical approach’ and suffers from a number of limitations. First, allelic variation in each cross is usually restricted because typically only two parents are used to create a QTL mapping population. Second, since early generation crosses are used, the number of recombination events per chromosome is usually small. Third, a typical QTL identified from a cross consisting of a few hundred offspring can span anywhere between a few to tens of centimorgans encompassing several megabases. Such large genome regions contain, typically, hundreds if not thousands of genes,

Table 2.1 QTL mapping studies in three important crop species

Species	Trait	QTLs	Population	Linkage groups covered	Markers linked to QTL	Software used	References
Rice	Root penetration ability	4 QTLs—number of penetrated roots 19 QTLs—total root number 6 QTLs—root penetration index 10 QTLs—tiller number	RIL ‘Indica (CO39) × Japonica (Moroberekan)’	Chr1, Chr9, Chr11, Chr12	RFLP	MAPMAKER/QTL	Ray et al. (1996)
	Submergence tolerance	Major gene, 4 QTLs	RIL ‘Indica (IR74 × FR13A)’	Major gene on Chr9, other QTLs on Chr6, Chr7, Chr11, Chr12	RFLP, AFLP	MAPMAKER/QTL	Nandi et al. (1997)
	Phosphorus deficiency resistance	Major QTLs	RIL ‘IR20 × IR55178-3B-9-3’	Chr12	AFLP, RFLP	MAPMAKER	Ni et al. (1998)
	Blast resistance	3 QTLs	F ₄ ‘Nipponbare × Owarihatamochi’	Chr4, Chr9, Chr12	RFLP	MAPMAKER	Fukuoka and Okuno (2001)
	Milling quality, protein content, and color characteristics	2 major QTLs	RIL ‘Zhenshan 97 × Minghui 63’	Chr3, Chr5	RFLP, SSRs	MAPMAKER/QTL	Tan et al. (2001)
	Drought tolerance	77 QTLs	DH ‘Japonica (CT9903 × Indica (IR62266))’	Chr3, Chr4, Chr8	RFLP, AFLP, SSRs	MAPMAKER, MQTL	Lanceras et al. (2004)
	Basal root thickness	Major QTLs	RIL ‘IRAT109 × Yuefu’	Chr4	SSRs	MAPMAKER/EXP, QTLMAPPER	Liu et al. (2008)
	Grain quality	16 QTLs	RIL ‘Indica (Chuan7) × Japonica (Nanyangzhan)’	Chr1, Chr2, Chr3, Chr6, Chr7, Chr9	SSRs	MAPMAKER, QTLMapper	Lou et al. (2009)
	Drought resistance	22 QTLs	RIL ‘Indica (IR20 × Nootripathu)’	Chr1, Chr2, Chr4, Chr5, Chr6, 11	SSRs	MAPMAKER/EXP MS-DOS, WINQTL-CART	Gomez et al. (2010)
	Leaf width and grain number	2 major QTLs	F ₂ ‘Indica (HP) × Nipponbare’	Chr1	SSRs, Indel	QTL IciMapping v 3.1	Tian et al. (2014)

(continued)

Table 2.1 (continued)

Species	Trait	QTLs	Population	Linkage groups covered	Markers linked to QTL	Software used	References
Wheat	Drought-induced abscisic acid production	1 QTL	DH 'Chinese spring $\times$ SQ1'	Chr 5AL	RFLP	MAPMAKER, MAPMAKER-QTL	Quarrie et al. (1994)
	Preharvest sprouting tolerance	One major QTL	RIL 'SPR8 198 $\times$ HD2329'	Chr 3A	SSRs, AFLP, SAPML	MAPMAKER/EXP, QTL cartographer	Kulwal et al. (2005)
	Growth response to exogenously applied stress-induced hormones	9 QTLs	DHR 'Chinese Spring (CS) $\times$ Synthetic 6A (S6A)'	Chr 6AS	SSRs	QGENE software	Castro et al. (2008)
	Stripe rust resistance	2 QTLs	RIL 'Stephens $\times$ Michigan Amber'	Chr 6BS	SSRs	Mapmaker MacIntosh, Win QTL Cart software	Santra et al. (2008)
	Aluminum toxicity	2 major QTLs	'CS7'/CS (Synthetic 3B)' DH lines	Chr 4D, Chr 3B	SSRs	MAPMAKER, QGENE software	Navakode et al. (2009)
	Senescence-related traits under high temperature	16 QTLs	RIL 'Vennor $\times$ Karl 92'	Chr2A, Chr 6A, Chr 6B, Chr3A, Chr 3B, Chr 7A	SSRs, AFLP	MAPMAKER/EXP, QTL cartographer	Vijayalakshmi et al. (2010)
	Terminal heat tolerance	3 major QTLs	RIL 'NW1014XHUIW468'	Chr2B, Chr7B, Chr7D	SSRs	MAPMAKER, QTL cartographer	Paliwal et al. (2012)
	Grain yield under multi-environment	One major QTL	DH 'RAC875 $\times$ Kukri' RIL 'RAC875 $\times$ Kukri'	Chr 3B	SSRs, ISBP	MapManager Version QTXb20, MET QTL model	Bonneau et al. (2013)
	Powdery mildew resistance	4 QTLs	DH 'Pingyuan 50 $\times$ Mingxian169'	Chr2BS, Chr3BS, Chr5AL, Chr3BL	SSRs	MapManager QTXb20, QTL cartographer	Asad et al. (2014)
	Root penetration ability	13 QTLs	DH 'Cranbrook $\times$ Halberd'	Chr 2D, Chr 4A, Chr 6B, Chr 7B	AFLP, DArT	MultiQTL	Acuña et al. (2014)

(continued)

Table 2.1 (continued)

Species	Trait	QTLs	Population	Linkage groups covered	Markers linked to QTL	Software used	References
Maize	Drought tolerance	5 QTLs	SD34 × SD35 F3 families	Chr1, Chr3, Chr5, Chr6, Chr8	RFLP	Mapmaker, Mapmaker/QTL	Agrama and Moussa (1996)
	Downy mildew resistance	3 QTLs	RIL 'G62 × G58'	Chr1, Chr9	RFLP	MAPMAKER/EXP, single factor analysis for QTL identification	Agrama et al. (1999)
	Early plant vigor in chilly environments	17 QTLs	DH lines 'SL and TH'	Chr2, Chr3, Chr4, Chr5, Chr6, Chr7, Chr8, Chr10	SSRs	PL/ABQTL	Presteri et al. (2007)
	Head smut resistance	5 QTLs	F _{2:3} families 'Mo17 × Huangzao4'	Chr1, Chr2, Chr3, Chr8	SSRs, AFLP	Mapmaker/EXP, Winqtl cartographer	Li et al. (2008)
	Reaction to phaeosphaeria leaf spot	6 QTLs	F ₂ 'L14-04B × L08-05F'	Chr1, Chr3, Chr4, Chr6, Chr8	SSRs	MAPMAKER/EXP, multiple interval mapping to map QTL	Moreira et al. (2009)
	Leaf temperature responses to drought	9 QTLs	RIL 'Zong3 × 87-1'	Chr1, Chr2, Chr9, Chr10	SSRs	QTL cartographer	Liu et al. (2011)
	Stay green traits	3 major QTLs; total 22 QTLs	F ₂ 'A150-3-2 × Mo17'	Chr1, Chr2, Chr3, Chr4, Chr5, Chr6, Chr9	SSRs	MAPMAKER/EXP, QTL mapping by QTLNetwork-2.0 using mixed model based composite interval mapping	Wang et al. (2012a, b)
	Gray leaf spot resistance	4 QTLs	Inbred line 'Y32 × Q11'	Chr1, Chr2, Chr5, Chr8	SSRs	Mapmaker, QTL cartographer	Zhang et al. (2012)
	Root architectural traits	15 QTLs	RIL 'Oh43 × W64a (OhW), Ny821 × H99 (NyH)	Chr1, Chr2, Chr3, Chr4, Chr5, Chr7, Chr8, Chr9	SNPs	QTL package of R	Burton et al. (2014)
	Yield and agronomic traits	14 QTLs	F ₂ '02S6140 × KSS22'	Chr1, Chr2, Chr3, Chr7, Chr8, Chr10	SSRs	Mapmaker, QTL mapper	Park et al. (2014)

making the process of identifying the causal gene in a QTL region a tedious and quite time-consuming task through map-based cloning (Price 2006). In addition, for many organisms the generation of mapping populations is either not possible or at least very time-consuming. For instance, the long generation time of most forest trees have completely prevented any progress in elucidating the genetic basis of complex traits using QTL mapping experiments (Neale and Savolainen 2004).

These limitations, however, can be reduced with the use of ‘association mapping’ (Zhu et al. 2008). Turning the gene-tagging efforts from biparental crosses to natural populations (or germplasm collections) and from traditional QTL mapping to linkage disequilibrium (LD)-based association mapping can lead to the most effective utilization of ex situ conserved natural genetic diversity of worldwide crop germplasm resources. This approach has many major advantages over conventional QTL mapping. First, a much larger and more representative gene pool can be surveyed. Second, it bypasses the expense and time of mapping studies and enables the mapping of many traits in one set of genotypes. Third, a much finer mapping resolution can be achieved, resulting in small confidence intervals of the detected loci compared to classical mapping, where the identified loci need to be fine-mapped. Finally, it has the potential not only to identify and map QTLs but also to identify the causal polymorphism within a gene that is responsible for the difference in two alternative phenotypes.

2.5.1 General Steps in Association Mapping

The performance of association mapping includes the following general steps: (1) selection of a group of individuals from a natural population or germplasm collection with wide coverage of genetic diversity; (2) measuring the phenotypic characteristics (yield, quality, tolerance, resistance, etc.) in the population, preferably, in different environments and multiple replication/trial design; (3) genotyping the mapping population individuals with molecular markers; (4) quantification of the extent of LD of a chosen population genome using molecular marker data; (5) assessment of the population structure (the level of genetic differentiation among groups within a sampled population individuals) and kinship (coefficient of relatedness between pairs of each individuals within a sample); and (6) based on information gained through quantification of LD and population structure, correlation of phenotypic and genotypic data with the application of an appropriate statistical approach that reveals ‘marker tags’ positioned within close proximity of targeted trait of interest.

2.5.2 *Selection of Association Mapping Panels*

Various kinds of populations can be used for association mapping; gene bank collections, elite breeding materials and/or specialized populations (e.g., NAM or MAGIC). In the case of gene bank materials, core collections are expected to represent most of the genetic variability with a manageable number of accessions and thus are suitable for association studies. In addition, the process of selecting a minimum sample size with maximum variation has a normalizing effect that is expected to reduce population structure and decrease LD, thus creating a situation favorable for association mapping. Core collections are useful for mapping qualitative traits, such as disease resistance or quality characteristics. Their broad genetic variability makes them often unsuitable for analysis of quantitative traits because accessions are usually unadapted to growing conditions, resulting in poor precision of trait measurement.

In plant breeding programs, a large body of phenotypic data is accumulated for elite lines from replicated field experiments over locations and years, thereby saving time to develop a panel. The use of those data for association mapping requires statistical models accounting for covariances introduced both by experimental design (years, locations, replicates) and polygenic effects. Moreover, those data are often unbalanced because new lines are included in field trials each year, while other lines are discarded. Population structure and higher LD can be prominent in elite material because it is common for closely related lines to be admitted to advanced trials. However, if pedigrees are known, the relationships among the lines can be determined and used to control for polygenic effects. Although association mapping in elite lines may not offer much improved resolution compared with QTL analysis in biparental mapping populations, there are at least two important advantages: a substantially higher level of polymorphism and detection of favorable alleles directly in the target population. Elite lines might be more desirable materials for mapping low-heritability traits such as grain yield, as the material is genetically more stable and are well adapted to normal growing conditions.

To increase the power and mapping resolution of marker–trait associations, some specialized populations have been constructed utilizing a joint strength of QTL mapping and AM. For example, nested association mapping (NAM) populations and multiparent advanced generation intercross (MAGIC) populations have been developed in many crops (Kover et al. 2009; McMullen et al. 2009; Huang et al. 2012). NAM populations are developed by crossing a set of diverse lines (up to 25) to one reference line. F_1 s of each cross are then selfed to develop RILs for each population. MAGIC populations are created by several generations of intercrossing among multiple founder lines, for example, four or eight lines. Multiple founders similar to a NAM population capture more allelic diversity than biparental mapping populations, whereas the multiple cycles of intercrossing give greater opportunity of recombination and hence greater precision of QTL mapping. However, it should be kept in mind that generating such specialized populations requires a lot of effort, time, and investment.

2.5.3 *Concept and Calculation of Linkage Disequilibrium (LD)*

The terms linkage and LD are also often confused. Linkage refers to the correlated inheritance of loci through the physical connection on a chromosome, whereas LD refers to the correlation between alleles in a population (Flint-Garcia et al. 2003) but not necessarily on the same chromosome. As a starting point for association mapping, it is important to gain knowledge of the patterns of LD for genomic regions of the ‘target’ organisms and the specificity of the extent of LD among different populations or groups to design and conduct unbiased association mapping. The two most commonly used statistics to measure LD are r^2 (square of the correlation coefficient) and D' (disequilibrium coefficient). The statistics r^2 and D' reflect different aspects of LD and perform differently under various conditions. The r^2 is affected by both mutation and recombination, while D' is affected by more mutational histories.

There are freely available softwares such as GOLD (Abecasis and Cookson 2000), TASSEL (www.maizegenetics.net), and Powermarker (Liu and Muse 2005) to depict the structure and pattern of LD. One can estimate an average genome-wide decay of LD by plotting LD values (r^2 values) obtained from a data set covering an entire genome (i.e., with more or less evenly spaced markers across all chromosomes in a genome) against distance. When such a LD decay plot generated, the usual practice is to look for distance point where LD value (r^2) decreases below 0.1 or half strength of D' ($D' = 0.5$) based on the curve of nonlinear logarithmic trend line. This gives the rough estimates of the extent of LD for association study, but for more accurate estimates, highly significant threshold LD values ($r^2 \geq 0.2$) are also used as a cutoff point. The decrease of the LD within the genetic distance indicates that the portion of LD is conserved with linkage and proportional to recombination (Gupta et al. 2005).

2.5.4 *Types of Association Mapping*

Association mapping generally falls into two broad categories: (1) candidate-gene-based association mapping, which relates polymorphisms in selected candidate genes that have purported roles in controlling phenotypic variation for specific traits and (2) genome-wide association mapping (GWAM), which surveys genetic variation in the whole genome to find signals of association for various complex traits (Risch and Merikangas 1996; Zhu et al. 2008). The absolutely most important aspect when deciding between a candidate gene approach and a whole-genome study is the extent of LD (see section on LD below) in the organism of interest, because the extent of LD determines not only the mapping resolution that can be achieved, but also the numbers of markers that are needed for an adequate coverage of the genome in a genome-wide study (Whitt and Buckler 2003). In species where

LD extends over long physical distances, relatively few markers are needed to ensure adequate genome coverage. For example, in species like *Arabidopsis thaliana* or in inbred lines of barley, LD can extend for tens or even hundreds of kilo base pairs, thus allowing GWAM with a relatively low number of evenly spaced markers (Aranzana et al. 2005; Rostoks et al. 2006). However, in many predominantly or obligately outcrossing organisms, such as maize (Remington et al. 2001) and many forest trees (Neale and Ingvarsson 2008), LD only extends a few hundred base pairs. For these species, several million markers would have to be genotyped to ensure adequate genome-wide coverage. Hence, in such species before several million markers (mainly single nucleotide polymorphisms) are generated, candidate-gene-based association mapping can be equally useful.

A candidate-gene-based association study is more hypothesis-driven than a genome-wide study because mapping is restricted to genes thought to be good candidates for controlling the trait of interest (Neale and Savolainen 2004; Hall et al. 2010). Although the selection of candidate genes is not always straightforward, genes are usually selected based on information obtained from, for instance, genetic, biochemical, mutation, physiology, or expression studies in both model and non-model plant species. The construction of molecular linkage maps based on the genes [e.g., expressed sequence tags (ESTs), EST-SSRs]] is another way of identifying the candidate genes underlying QTL, instead of time-consuming fine mapping (Sehgal and Yadav 2009). Standard neutrality tests applied to DNA sequence variation data can also be used to select candidate genes or amino acid sites that are putatively under selection for association mapping. This is one of the effective alternative strategies in association mapping that allow reducing the total amount of marker genotyping in less number of individuals. This increases the power and precision of the trait–marker correlations. However, it is important to remember that a candidate gene approach is limited by the choice of candidate genes that are identified and hence always runs the risk of missing out on identifying causal mutations located in non-identified candidate genes.

2.5.5 Candidate-Gene-based Association Studies in Plants

The candidate gene strategy has shown promise for bridging the gap between quantitative genetic and molecular genetic approaches to study complex traits (Cattivelli et al. 2008; Ingvarsson and Street 2011). Along this line, important studies on association mapping with the candidate gene approach are summarized as follows. Vernalization requirement in wheat is controlled by four major genes, viz. *VRN1*, *VRN2*, *VRN3*, and *VRN4*, with *VRN1* gene copies *VRN-A1*, *VRN-B1*, and *VRN-D1* located on the long arms of chromosomes 5A, 5B, and 5D, respectively (Yoshida et al. 2010). An association mapping study conducted by Rousset et al. (2011) on 235 hexaploid wheat collections revealed the effects of the flowering time candidate genes in modulating flowering time in wheat. In that study,

genetic variation in *VRN-A1*, *VRN-B1*, and *VRN-D1* genes explained a large part of phenotypic variation in growth habit. Huang and Brûlé-Babel (2012) studied genetic diversity, haplotype structure, and association of genes involved in starch biosynthesis in wheat. Genes encoding granule-bound starch synthase (GBSSI, also known as waxy or *Wx* genes) and soluble starch synthase (SSIIa) were selected for nucleotide diversity and SNP density study. None of the SNPs within the three *SSIIa* genes and *Wx-D1* gene was associated with yield-related traits. However, both SNPs and haplotypes within the *Wx-A1* gene were associated with seed number per spike, seed weight per spike and thousand kernel weight. Another study on grain size of wheat also demonstrated the association of haplotype of a grain size gene (*TaGW2*) with larger grain size, earlier heading date and maturity in hexaploid wheat (Su et al. 2011). Similarly, transcription factors such as the gibberellin-regulated Myb factor (GAMYB), the barley leucine zippers 1 and 2 (*BLZ1*, *BLZ2*), and the barley prolamin box binding factor (BPBF) were evaluated for their association with agronomic traits in barley (Haseneyer et al. 2010). SNPs within *BLZ1* were associated with days to flowering, and its haplotype was also associated with both days to flowering and plant height. The haplotype of *BLZ2* was associated with thousand kernel weight, while the haplotype of the *BPBF* gene was associated with both crude protein and starch in barley endosperm (Haseneyer et al. 2010). However, the candidate genes explained only a small portion of the total genetic variation. Similarly for sorghum and rice, candidate genes involved in starch biosynthesis were associated with the expected traits and the results were in agreement with QTL studies (Bao et al. 2006; Figueiredo et al. 2010).

More than 20 studies of candidate gene association analysis in maize have been published to date. These studies used candidate genes from well characterized and relatively simple metabolic pathways (Wilson et al. 2004; Harjes et al. 2008; Yan et al. 2010) or those with extensive prior evidence for the role of the candidate gene (s) in the control of the phenotype of interest such as map-based cloning studies (Salvi et al. 2007; Ducrocq et al. 2008; Zheng et al. 2008; Buckler et al. 2009). The candidate genes were also selected based on information in closely related species (Li et al. 2010a, b), and information from QTL mapping studies and/or expression results (Krill et al. 2010). The most comprehensive candidate gene association results have been recently reported in maize for SNPs identified from 540 genes putatively involved in accumulation of carbohydrate and ABA metabolites during drought stress (Setter et al. 2010). In this study, the SNP from a homologue of an *Arabidopsis* MADS-box gene was significantly associated with phaeic acid in ears of irrigated plants, while a SNP in pyruvate dehydrogenase kinase was significantly associated with silk sugar concentrations. Similarly, a SNP from an aldehyde oxidase gene was associated with ABA levels in silk under non-irrigated conditions.

The candidate gene association mapping approach has been widely applied in forest tree genetics studies as developing a biparental population is practically unfeasible for most conifers. González-Martínez et al. (2006) studied the pattern of polymorphisms of 18 drought-responsive candidate genes in 32 *Pinus taeda* L. individuals. LD within the sequenced gene regions varied from low to high

depending on the candidate gene locus. A total of 196 SNPs and 82 LD blocks were obtained in 18 candidate gene loci. By constructing LD blocks, 94 haplotype SNPs were identified to improve the LD values and were successfully used in detecting significant r^2 values for LD blocks study. The same authors evaluated the association of four candidate genes belonging to different functional classes with carbon isotope discrimination (CID) at two locations. The genes were general protection factor (*dhn-1*), antioxidants (*sod-chl*), transcription factor (*wrky-like*), and putative cell wall protein (*lp5-like*). Antioxidant (*sod-chl*) and *Cu/Zn superoxide dismutase* genes showed significant association with CID at both locations. However, none of the significant associations explained a substantial amount of phenotypic variance in CID.

2.5.6 *Confounding Effects of Population Structure in Association Mapping*

One of the main hurdles for using association mapping to dissect the genetic architecture of complex traits in plants is the risk of incurring false positives due to population structure (Pritchard et al. 2000). The problem of population structure arises because any phenotypic trait that is also correlated with the underlying population structure at neutral loci will show an inflated number of positive associations resulting in Type I errors. Among many methods developed to deal with this problem, ‘genomic control’ (GC) method (Devlin and Roeder 1999) estimates association using a large number of putative neutral markers or markers not thought to be involved in controlling the trait of interest. The distribution of the test statistic of interest is then calculated from these associations and a critical value corresponding to the desired Type I error rate is chosen from this distribution. Another method that is commonly used to control for population structure is structured associations (SA) (Pritchard et al. 2000). SA first searches a population for closely related clusters/subdivisions using Bayesian approach and then uses the clustering matrices (Q) in association mapping (by a logistic regression) to correct the false associations. Population structure and shared co-ancestry coefficients between individuals of subdivisions of a population can be effectively estimated with *STRUCTURE* program (Pritchard et al. 2000) using several models for linked and unlinked markers.

Principal component analysis (PCA) was recently suggested as a fast and effective way to diagnose population structure (Zhu and Yu 2009). The PCA method summarizes variation observed across all markers into a number of underlying component variables and these components, typically the first few, can then be used to replace Q to adjust for population structure. The PCA method makes it computationally feasible to handle a large number of markers (tens of thousands) and correct for subtle population stratification. There are many programs

that can be used to calculate PCA such as DARwin (Perrier and Jacquemoud-Collet 2006) and EIGENSTRAT (Price et al. 2006).

However, incorporating only population structure information in the analysis is not good enough itself when highly structured population with some degree of related individuals are used in the association mapping. A mixed linear model (MLM) that combines both population structure information (Q -matrix or PCA) and level of pairwise relatedness coefficients (kinship-matrix) should be used in the analysis. Although computationally intensive, MLM approach is very effective in removing the confounding effects of the population in association mapping (Yu et al. 2005).

2.5.7 Association with Raw Data or BLUPs or Residuals

In general, raw data can be used directly in association analysis provided it is available for all entries and for all replicates in different locations/years. For cases where phenotypes are not evaluated for all individuals and replicates due to large sample size, best linear unbiased predictors (BLUPs) from a mixed model may be substituted as the dependent variable. In such cases, the association analysis using BLUPs can be performed with many fewer observations and require much less time.

More recently, researchers have also used residuals instead of raw data. The rationale is that after removing all the effects except the marker, including the polygenic genetic variance captured by the BLUPs, the signal due to marker association is still contained in the residuals. Signal from the marker will be removed only to the extent that it is correlated with the other effects. The residual approach performs as well as the approach using raw phenotype directly for low-heritability traits (Aulchenko et al. 2007). Because the association test using residuals is performed without including the polygenic random effect, tests of individual markers run quickly. The mixed model equations with thousands of individuals only need to be solved once for any particular phenotype. After that, the millions of association tests for individual markers can then be performed using simple t -tests or F -tests of the marker classes.

2.5.8 Association Analysis Programs

Public, freely available software suitable for association analysis using mixed models in plants includes TASSEL and EMMA/R. Both analyze moderately large data sets in a reasonable amount of time but only allow a single effect (samples or taxa) to be fit as a random effect. All other effects are treated as fixed. EMMA relies on the R for data management and visualization, whereas TASSEL handles those functions itself. Several commercial software packages available for association

studies include ASREML, JMP Genomics, SAS, and GenStat. ASREML and JMP Genomics are specifically engineered for genetic analysis and can handle more complex models, whereas general purpose packages such as SAS Proc Mixed and GenStat can perform association analysis but require more expertise and programming on the part of the user.

From the user's perspective, clearly freely available software such as TASSEL plays an important role in scientific investigation. Another advantage of using TASSEL is that both graphical user interface (GUI) and command line interface (CLI) versions exist. In the GUI, the plug-ins are invoked by clicking buttons on the interface. With the CLI, the plug-ins are used in a predetermined pipeline that passes the output from one step to the input of another. Hence, scientists can use these versions depending on their expertise and consistent results are achieved independent of the interface. In the latest version of TASSEL (TASSEL 5.0), compressed MLM method is available for computing large data sets with up to 500,000 markers.

2.5.9 Significance Threshold

A threshold is set to declare significant associations. Any of the two statistical methods can be used to correct for multiple comparisons: false discovery rate (FDR) and Bonferroni correction. The correction is needed whenever one would like to test multiple hypotheses simultaneously. FDR controls the expected proportion of false positives among significant results by determining a threshold from the observed p value distribution in the data, whereas Bonferroni corrections control the chance of any false positives (Benjamini and Hochberg 1995). Given the aims of the study, one may consider a high FDR for some projects (e.g., investigating the genetic architecture of a trait) and a low FDR for others (e.g., identifying candidate loci for follow-up studies).

2.5.10 Validation of Association Results

Table 2.2 summarizes genome-wide association studies conducted in three major crops. It clearly shows the increase in association studies in wheat, maize, and rice for dissecting various complex traits. As the number of studies documenting alleles showing significant associations with quantitative trait variation, there is an increasing need to replicate findings before marker information is incorporated in selection decisions, or before large sums are invested into identification of causal factors and gene cloning. The most straightforward way is to compare the association mapping results with previous results published for the trait, for example, using biparental populations. If markers in close proximity (within 10 cM) to previously reported QTLs/genes are identified, the result will not only be validated

Table 2.2 Association mapping studies in three important crops

Crop	Marker system	Trait analyzed	Marker–trait association	Program used for genetic structure and association analysis	References
Rice	SSRs	Grain yield (GY), kernel length (KL), kernel width (KW), kernel length/width ratio (LWR), 1000 kernel weight (TKW)	GY-5 markers, KL-6 markers, KW-5 markers, LWR-5 markers, TKW-4 markers	STRUCTURE, PowerMarker 3.23, Arlequin 3.0, TASSEL	Agrana et al. (2007)
	SSRs, InDel	Heading date surveyed in 2006 (HD06), heading date in 2006 (HD07), plant height in 2006 (PH06), plant height in 2007 (PH07), panicle length in 2006 (PL06), panicle length in 2007 (PL07)	HD06-3 markers, HD07-3 markers, PH06-3 markers, PH07-2 markers, PL06-2 markers, PL07-1 marker	STRUCTURE, NYSYS-pc, SPAGeDi, TASSEL	Wen et al. (2009)
	SSRs, InDel	Single stigma exertion (SSigE), dual exertion (DSigE), total stigma exertion (TSigE), spikelet length (SpkL), spikelet width (SpkW), spikelet length/width ratio (SpkL/W)	SSigE-4 markers, DSigE-6 markers, TSigE-5 markers, SpkL-4 markers, SpkW-3 markers, SpkL/W-7 markers	STRUCTURE, GenAlex, SPAGeDi, TASSEL	Yan et al. (2009)
	SSRs	Grain yield (YLD), Tiller number per plant (TILN), Panicle number per plant (PANN), Yield from ratining (RYLD), Amylose content (AC), Head milled rice (MR)	Low-land accessions -PANN (2004)-1 marker, AC (2004, 2005)-1 marker, MR (2005)-2 markers, YLD (2005)-4 markers Upland accession -AC (2004, 2005)-1 marker	STRUCTURE, FSTAT 2.9.3.2, Genetix 4.03, SPAGeDi, TASSEL	Borba et al. (2010)
	SSRs	Sheath blight (ShB) resistance	ShB resistance-10 markers	STRUCTURE, PowerMarker, MEGA 4.0, NTSYSpc, TASSEL	Jia et al. (2012)
	SSRs, InDel	Harvest index traits in temperate (Arkansas) and subtropical climates (Texas)	Harvest index traits in Arkansas-36 markers, harvest index traits in Texas-28 markers	STRUCTURE, SPAGeDi, PDMIX800 SAS macro	Li et al. (2012)
	SSRs	Grain yield per plant, weight of 100 grains, panicle length	Grain yield per plant-3 markers, weight of 100 grains-1 marker, panicle length-5 markers	DARWin, STRUCTURE, TASSEL	Vanniarajan et al. (2012)
	SSRs	Cold tolerance	Cold tolerance-total 24 markers, in Yunnan (natural low temperature)-5 markers, in Jilin (cold water irrigation)-19 markers	STRUCTURE, Arlequin, SPAGeDi, NTSYSpc, TASSEL	Cui et al. (2013)
	SSRs	Seed vigor evaluated by root length (RL), shoot length (SL), shoot dry weight (SDW)	RL-7 markers, SL-12 markers, SDW-8 markers	STRUCTURE, Arlequin, SPAGeDi, TASSEL	Dang et al. (2014)
	SNPs	Genotype × Environment interactions for agronomic traits	10 agronomic traits-23 putative loci	STRUCTURE, TASSEL	Xu et al. (2014)

(continued)

Table 2.2 (continued)

Crop	Marker system	Trait analyzed	Marker–trait association	Program used for genetic structure and association analysis	References
Wheat	SSRs	Kernel size and milling quality	Kernel size and milling quality-62 loci	STRUCTURE, TASSEL	Bresegrello and Sorrells (2006)
	DAT	Resistance to leaf rust (LR), powdery mildew (PM), stripe rust (SR), and yellow (stripe) rust (YR), plus grain yield (GY)	LR-87 markers, PM-61 markers, SR-63 markers, YR-122 markers, GY-213 markers	STRUCTURE, DARWIN, TASSEL	Crossa et al. (2007)
	SSRs	Russian wheat aphid resistance	Leaf chlorosis-28 markers, Leaf rolling-8 markers	STRUCTURE, TASSEL	Peng et al. (2009)
	SSRs	Drought-adaptive traits and grain yield in different environment different for water availability	Grain yield-56 markers in one environment, 20 markers in two environments, 3 markers in 5 environments, only one marker in six environments	STRUCTURE, TASSEL	Maccaferri et al. (2010)
	SSRs	Plant height under well water and drought stress condition	Plant height and drought tolerance coefficient-23 markers	PowerMarker, STRUCTURE, TASSEL	Zhang et al. (2011)
	SSRs	Preharvest sprouting tolerance (PHST)	PHST-30 markers	STRUCTURE, TASSEL	Jaiswal et al. (2012)
	STS, DAT	Stem rust resistance	Stem rust resistance-12 markers	SAS PROC PRINCOMP, TASSEL	Yu et al. (2012)
	DAT	Pest resistance—hessian fly (HF), Russian wheat aphid (RWA), sunn pest (SP), wheat stem saw fly (WSSF), and cereal leaf beetle (CLB)	HF-13 markers, RWA-13 markers, SP-17 markers, WSSF-5 markers, CLB-5 markers	STRUCTURE, TASSEL	Joukhadar et al. (2013)
	SNPs	Soil-borne wheat mosaic virus (SBWMV) resistance	SBWMV resistance-6 markers	STRUCTURE, TASSEL	Liu et al. (2014)
	SNPs	Stripe rust caused by (<i>Puccinia striiformis</i> f. sp. <i>tritici</i> ; <i>Pst</i>), resistance	Resistance to <i>pst</i> at seedling stage-27 markers, resistance to <i>pst</i> at adult plant stage-38 markers	STRUCTURE, TASSEL	Zegeye et al. (2014)

(continued)

Table 2.2 (continued)

Crop	Marker system	Trait analyzed	Marker–trait association	Program used for genetic structure and association analysis	References
Maize	SNPs	Plant height	Plant height-204 markers	PLINK, STRUCTURE, TASSEL	Weng et al. (2011)
	SSRs	Downy mildew resistance	Downy mildew resistance-3 markers	STRUCTURE, TASSEL	Phumichai et al. (2012)
	SSRs, SNPs	Flowering time (FT), northern corn leaf blight resistance (NCLB), NCLB resistance corrected for FT	FT-7 markers, NCLB-4 markers, NCLB resistance corrected for FT-4 markers	STRUCTURE, ASREML release, EMMA	Van Ingelhardt et al. (2012)
	SNPs, InDels	Head smut resistance	Head smut resistance-19 markers and 18 candidate genes	STRUCTURE, SPAGeDi, NTSYSpc, TASSEL	Wang et al. (2012a, b)
	SNPs	Chilling tolerance	Chilling tolerance-43 markers	STRUCTURE, SPAGeDi, NTSYSpc, TASSEL	Huang et al. (2013)
	SNPs	Oil biosynthesis in maize kernels	Kernel oil concentration and fatty acid composition-26 SNPs and candidate genes associated with oil concentration	TASSEL	Li et al. (2013)
	SNPs	Agronomic traits under well watered and water stresses conditions	126 trait $\times$ environment $\times$ treatment combinations-42 markers	STRUCTURE, TASSEL	Xue et al. (2013)
	SNPs	Waterlogging tolerance	Traits under normal and waterlogged condition (seedling height, root length, shoot fresh weight, root fresh weight, root dry weight, root dry weight)-47 markers	STRUCTURE, SPAGeDi, NTSYSpc, TASSEL	Zhang et al. (2013)
	SNPs	Gray leaf spot (GLS) resistance	GLS resistance-51 markers	STRUCTURE, TASSEL	Shi et al. (2014)
	SNPs	Agronomic traits in an enlarged maize association panel	Ear leaf width, Ear length, kernel width, plant height and tassel main axis length-19 markers	TASSEL	Yang et al. (2014)

but also increase the confidence to pursue the new genomic target identified for the trait. Secondly, results can be validated in different populations. This is more reliable as the probability of observing false positives becomes small if significant associations are confirmed in two or more validation populations. Third, if association studies point to alleles with opposite effects on a trait of interest, one can generate multiple F_2 populations from parents that harbor contrasting alleles and determine whether differences in phenotype co-segregate with the locus in question. Once markers tightly linked to the target trait are validated, they provide several magnitudes of return on investment through increased speed and cost efficiency of breeding programs.

However, failure to replicate a previously documented association can occur because of a large number of issues, both in the initial and the replication study, including factors such as difficulties in replicating the environment, small sample size, poor study design, or lack of rigorous phenotype scoring (Manolio et al. 2009). Another concern is that allelic effects of previously documented associations usually decline in replication studies. This phenomenon is known as the ‘Beavis effect’ (Beavis 1994) in the QTL mapping literature and occurs because significant associations are reported only when test statistics exceed a predetermined critical threshold. The estimated effects of the detected associations are therefore sampled from a truncated distribution, and the weaker the initial effect the more serious this overestimation is (Rockman 2008). The Beavis effect has also been shown to occur in association mapping studies (Ingvarsson et al. 2008). Hence, careful consideration of the power of the prospective association study should be taken early on in the experiment, so that things like the Beavis effect can be minimized or eliminated.

2.6 Conclusions

Comparison of linkage analysis and association mapping for QTL detection revealed that linkage mapping is more useful for genome-wide scan for QTLs, while association mapping gives more precise location of an individual QTL. Therefore, linkage analysis may be preferred for preliminary location of QTLs and then use association mapping for more precise location. Association mapping is prone to the identification of false positives, especially if the experimental design is not rigorously controlled. For example, population structure has long been known to induce many false positives and accounting for population structure has become one of the main issues when implementing association mapping in plants. Also, with increasing numbers of genetic markers used, the problem becomes separating true from false positive and this highlights the need for independent validation of identified association. The examples of association mapping studies performed in three most important crops’ germplasm largely demonstrate the flourish of crop genomics era with the utilization of powerful LD-based association mapping tool. Currently, a number of such studies are in progress for various other crops in many laboratories worldwide. The near-future completion of genome sequencing projects

of crop species, powered with more cost-effective sequencing technologies, will certainly create a basis for application of whole-genome-association studies. This will provide with more powerful association mapping tool(s) for crop breeding and genomics programs in tagging true functional associations conditioning genetic diversities, and consequently, its effective utilization.

References

- Abecasis GR, Cookson W (2000) GOLD—graphical overview of linkage disequilibrium. *Bioinformatics* 16:182–183
- Acuña TB, Rebetzke G, He X et al (2014) Mapping quantitative trait loci associated with root penetration ability of wheat in contrasting environments. *Mol Breed* 1–12
- Agrama HA, Moussa ME (1996) Mapping QTLs in breeding for drought tolerance in maize (*Zea mays* L.). *Euphytica* 91:89–97
- Agrama H, Moussa M, Naser M et al (1999) Mapping of QTL for downy mildew resistance in maize. *Theor Appl Genet* 99:519–523
- Agrama H, Eizenga G, Yan W (2007) Association mapping of yield and its components in rice cultivars. *Mol Breed* 19:341–356
- Aranzana MJ, Kim S, Zhao K et al (2005) Genome-wide association mapping in Arabidopsis identifies previously known flowering time and pathogen resistance genes. *PLoS Genet* 1:e60
- Asad MA, Bai B, Lan C et al (2014) Identification of QTL for adult-plant resistance to powdery mildew in Chinese wheat landrace Pingyuan 50. *Crop J* 2:308–314
- Asins M (2002) Present and future of quantitative trait locus analysis in plant breeding. *Plant Breed* 121:281–291
- Aulchenko YS, De Koning D-J, Haley C (2007) Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. *Genetics* 177:577–585
- Bao J, Corke H, Sun M (2006) Microsatellites, single nucleotide polymorphisms and a sequence tagged site in starch-synthesizing genes in relation to starch physicochemical properties in nonwaxy rice (*Oryza sativa* L.). *Theor Appl Genet* 113:1185–1196
- Beavis WD (1994) The power and deceit of QTL experiments: lessons from comparative QTL studies. In: *Proceedings of the forty-ninth annual corn and sorghum industry research conference* pp 250–266
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc Ser B (Methodol)* 289–300
- Bonneau J, Taylor J, Parent B et al (2013) Multi-environment analysis and improved mapping of a yield-related QTL on chromosome 3B of wheat. *Theor Appl Genet* 126:747–761
- Borba TCDO, Brondani RPV, Breseghello F et al (2010) Association mapping for yield and grain quality traits in rice (*Oryza sativa* L.). *Genet Mol Biol* 33:515–524
- Breseghello F, Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics* 172:1165–1177
- Buckler ES, Holland JB, Bradbury PJ et al (2009) The genetic architecture of maize flowering time. *Science* 325:714–718
- Burton AL, Johnson JM, Foerster JM et al (2014) QTL mapping and phenotypic variation for root architectural traits in maize (*Zea mays* L.). *Theor Appl Genet* 1–19
- Castro A, Tacaliti M, Giménez D et al (2008) Mapping quantitative trait loci for growth responses to exogenously applied stress induced hormones in wheat. *Euphytica* 164:719–727
- Cattivelli L, Rizza F, Badeck F-W et al (2008) Drought tolerance improvement in crop plants: an integrated view from breeding to genomics. *Field Crops Res* 105:1–14

- Cheema J, Ellis TH, Dicks J (2010) THREaD mapper studio: a novel, visual web server for the estimation of genetic linkage maps. *Nucl Acids Res* 38:188–193
- Collard B, Jahufer M, Brouwer J et al (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. *Euphytica* 142:169–196
- Crossa J, Burgueno J, Dreisigacker S et al (2007) Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. *Genetics* 177:1889–1913
- Cui D, Xu C-Y, Tang C-F et al (2013) Genetic structure and association mapping of cold tolerance in improved japonica rice germplasm at the booting stage. *Euphytica* 193:369–382
- Dang X, Thi TGT, Dong G et al (2014) Genetic diversity and association mapping of seed vigor in rice (*Oryza sativa* L.). *Planta* 239:1309–1319
- De Alencar Figueiredo L, Sine B, Chantereau J et al (2010) Variability of grain quality in sorghum: association with polymorphism in Sh2, Bt2, SssI, Ae1, Wx and O₂. *Theor Appl Genet* 121:1171–1185
- Devlin B, Roeder K (1999) Genomic control for association studies. *Biometrics* 55:997–1004
- Ducrocq S, Madur D, Veyrieras J-B et al (2008) Key impact of Vgt1 on flowering time adaptation in maize: evidence from association mapping and ecogeographical information. *Genetics* 178:2433–2437
- Faris JD, Haen KM, Gill BS (2000) Saturation mapping of a gene-rich recombination hot spot region in wheat. *Genetics* 154:823–835
- Flint-Garcia SA, Thornsberry JM, Iv B (2003) Structure of linkage disequilibrium in plants*. *Annu Rev Plant Biol* 54:357–374
- Fukuoka S, Okuno K (2001) QTL analysis and mapping of pi21, a recessive gene for field resistance to rice blast in Japanese upland rice. *Theor Appl Genet* 103:185–190
- Gomez SM, Boopathi NM, Kumar SS et al (2010) Molecular mapping and location of QTLs for drought-resistance traits in indica rice (*Oryza sativa* L.) lines adapted to target environments. *Acta Physiol Plant* 32:355–364
- González-Martínez SC, Ersoz E, Brown GR et al (2006) DNA sequence variation and selection of tag single-nucleotide polymorphisms at candidate genes for drought-stress response in *Pinus taeda* L. *Genetics* 172:1915–1926
- Gupta PK, Rustgi S, Kulwal PL (2005) Linkage disequilibrium and association studies in higher plants: present status and future prospects. *Plant Mol Biol* 57:461–485
- Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* 69:315–324
- Hall D, Tegström C, Ingvarsson PK (2010) Using association mapping to dissect the genetic basis of complex traits in plants. *Briefings in functional genomics* elp048
- Harjes CE, Rocheford TR, Bai L et al (2008) Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. *Science* 319:330–333
- Haseneyer G, Stracke S, Piepho H-P et al (2010) DNA polymorphisms and haplotype patterns of transcription factors involved in barley endosperm development are associated with key agronomic traits. *BMC Plant Biol* 10:5
- Hoisington D, Khairallah M, Reeves T et al (1999) Plant genetic resources: what can they contribute toward increased crop productivity? *Proc Natl Acad Sci* 96:5937–5943
- Huang BE, George AW, Forrest KL, Kilian A, Hayden MJ, Morell MK, Cavanagh CR (2012) A multiparent advanced generation inter-cross population for genetic analysis in wheat. *Plant Biotechnol J*. 2012 0(7):826-39
- Huang X-Q, Brûlé-Babel A (2012) Sequence diversity, haplotype analysis, association mapping and functional marker development in the waxy and starch synthase IIa genes for grain-yield-related traits in hexaploid wheat (*Triticum aestivum* L.). *Mol Breed* 30:627–645
- Huang J, Zhang J, Li W et al (2013) Genome-wide association analysis of ten chilling tolerance indices at the germination and seedling stages in maize. *J Integr Plant Biol* 55:735–744
- Ingvarsson PK, Street NR (2011) Association genetics of complex traits in plants. *New Phytol* 189:909–922

- Ingvarsson PK, Garcia MV, Luquez V et al (2008) Nucleotide polymorphism and phenotypic associations within and around the phytochrome B2 locus in European aspen (*Populus tremula*, Salicaceae). *Genetics* 178:2217–2226
- Jaiswal V, Mir R, Mohan A et al (2012) Association mapping for pre-harvest sprouting tolerance in common wheat (*Triticum aestivum* L.). *Euphytica* 188:89–102
- Jansen RC, Stam P (1994) High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* 136:1447–1455
- Jia L, Yan W, Zhu C et al (2012) Allelic analysis of sheath blight resistance with association mapping in rice. *PLoS ONE* 7:e32703
- Joukhadar R, El-Bouhssini M, Jighly A et al (2013) Genome-wide association mapping for five major pest resistances in wheat. *Mol Breed* 32:943–960
- Kover PX, Valdar W, Trakalo J et al (2009) A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genet* 5:e1000551
- Krill AM, Kirst M, Kochian LV et al (2010) Association and linkage analysis of aluminum tolerance genes in maize. *PLoS ONE* 5:e9958
- Kulwal P, Kumar N, Gaur A et al (2005) Mapping of a major QTL for pre-harvest sprouting tolerance on chromosome 3A in bread wheat. *Theor Appl Genet* 111:1052–1059
- Künzel G, Korzun L, Meister A (2000) Cytologically integrated physical restriction fragment length polymorphism maps for the barley genome based on translocation breakpoints. *Genetics* 154:397–412
- Lanceras JC, Pantuwan G, Jongdee B et al (2004) Quantitative trait loci associated with drought tolerance at reproductive stage in rice. *Plant Physiol* 135:384–399
- Lander ES, Green P, Abrahamson J et al (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1:174–181
- Li X, Wang Z, Gao S et al (2008) Analysis of QTL for resistance to head smut (*Sporisorium reilianum*) in maize. *Field crops Res* 106:148–155
- Li Q, Li L, Yang X et al (2010a) Relationship, evolutionary fate and function of two maize co-orthologs of rice GW2 associated with kernel size and weight. *BMC Plant Biol* 10:143
- Li Q, Yang X, Bai G et al (2010b) Cloning and characterization of a putative GS3 ortholog involved in maize kernel development. *Theor Appl Genet* 120:753–763
- Li X, Yan W, Agrama H et al (2012) Unraveling the complex trait of harvest index with association mapping in rice (*Oryza sativa* L.). *PLoS ONE* 7:e29350
- Li H, Peng Z, Yang X et al (2013) Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nat Genet* 45:43–50
- Lincoln S, Daly M, Lander E (1993) Constructing genetic linkage maps with MAPMAKER/EXP. Version 3.0. Whitehead Institute for Biomedical Research Technical Report, 3rd edn
- Liu K, Muse SV (2005) Powermarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21:2128–2129
- Liu L, Mu P, Li X et al (2008) Localization of QTL for basal root thickness in japonica rice and effect of marker-assisted selection for a major QTL. *Euphytica* 164:729–737
- Liu Y, Subhash C, Yan J et al (2011) Maize leaf temperature responses to drought: thermal imaging and quantitative trait loci (QTL) mapping. *Environ Exp Bot* 71:158–165
- Liu S, Yang X, Zhang D et al (2014) Genome-wide association analysis identified SNPs closely linked to a gene resistant to Soil-borne wheat mosaic virus. *Theor Appl Genet* 127:1039–1047
- Lou J, Chen L, Yue G et al (2009) QTL mapping of grain quality traits in rice. *J Cereal Sci* 50:145–151
- Ma X-F, Ross K, Gustafson J (2001) Physical mapping of restriction fragment length polymorphism (RFLP) markers in homoeologous groups 1 and 3 chromosomes of wheat by in situ hybridization. *Genome* 44:401–412
- Maccaferri M, Sanguineti MC, Demontis A et al (2010) Association mapping in durum wheat grown across a broad range of water regimes. *J Exp Bot* 91:287
- Manly KF, Cudmore RH Jr, Meer JM (2001) Map manager QTX, cross-platform software for genetic mapping. *Mamm Genome* 12:930–932

- Manolio TA, Collins FS, Cox NJ et al (2009) Finding the missing heritability of complex diseases. *Nature* 461:747–753
- Mcmullen MD, Kresovich S, Villeda HS et al (2009) Genetic properties of the maize nested association mapping population. *Science* 325:737–740
- Moreira JUV, Bento DAV, De Souza AP et al (2009) QTL mapping for reaction to phaeosphaeria leaf spot in a tropical maize population. *Theor Appl Genet* 119:1361–1369
- Nandi S, Subudhi P, Senadhira D et al (1997) Mapping QTLs for submergence tolerance in rice by AFLP analysis and selective genotyping. *Mol Gen Genet* MGG 255:1–8
- Navakode S, Weidner A, Lohwasser U et al (2009) Molecular mapping of quantitative trait loci (QTLs) controlling aluminium tolerance in bread wheat. *Euphytica* 166:283–290
- Neale DB, Ingvarsson PK (2008) Population, quantitative and comparative genomics of adaptation in forest trees. *Curr Opin Plant Biol* 11:149–155
- Neale DB, Savolainen O (2004) Association genetics of complex traits in conifers. *Trends Plant Sci* 9:325–330
- Ni J, Wu P, Senadhira D et al (1998) Mapping QTLs for phosphorus deficiency tolerance in rice (*Oryza sativa* L.). *Theor Appl Genet* 97:1361–1369
- Paliwal R, Röder MS, Kumar U et al (2012) QTL mapping of terminal heat tolerance in hexaploid wheat (*T. aestivum* L.). *Theor Appl Genet* 125:561–575
- Park KJ, Sa KJ, Kim BW et al (2014) Genetic mapping and QTL analysis for yield and agronomic traits with an F2: 3 population derived from a waxy corn × sweet corn cross. *Genes Genom* 36:179–189
- Paterson AH (1996) Making genetic maps. *Genome mapping in plants (biotechnology intelligence unit)* 23–39
- Peng J, Bai Y, Haley S et al (2009) Microsatellite-based molecular diversity of bread wheat germplasm and association mapping of wheat resistance to the Russian wheat aphid. *Genetica* 135:95–122
- Perrier X, Jacquemoud-Collet J (2006) DARwin software
- Phumichai C, Chunwongse J, Jampatong S et al (2012) Detection and integration of gene mapping of downy mildew resistance in maize inbred lines through linkage and association. *Euphytica* 187:369–379
- Prestrel T, Ouzunova M, Schmidt W et al (2007) Quantitative trait loci for early plant vigour of maize grown in chilly environments. *Theor Appl Genet* 114:1059–1070
- Price AH (2006) Believe it or not, QTLs are accurate! *Trends in Plant Science* 11:213–216
- Price AL, Patterson NJ, Plenge RM et al (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38:904–909
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Quarrie S, Gulli M, Calestani C et al (1994) Location of a gene regulating drought-induced abscisic acid production on the long arm of chromosome 5A of wheat. *Theor Appl Genet* 89:794–800
- Ray J, Yu L, Mccouch S et al (1996) Mapping quantitative trait loci associated with root penetration ability in rice (*Oryza sativa* L.). *Theor Appl Genet* 92:627–636
- Remington DL, Thornsberry JM, Matsuoka Y et al (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc Natl Acad Sci* 98:11479–11484
- Risch N (1992) Genetic linkage: interpreting LOD scores. *Science* 255:803–804
- Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. *Science* 273:1516–1517
- Rockman MV (2008) Reverse engineering the genotype–phenotype map with natural genetic variation. *Nature* 456:738–744
- Rostoks N, Ramsay L, Mackenzie K et al (2006) Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. *Proc Natl Acad Sci* 103:18656–18661

- Rousset M, Bonnin I, Remoué C et al (2011) Deciphering the genetics of flowering time by an association study on candidate genes in bread wheat (*Triticum aestivum* L.). *Theor Appl Genet* 123:907–926
- Salvi S, Sponza G, Morgante M et al (2007) Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proc Natl Acad Sci* 104:11376–11381
- Santra D, Chen X, Santra M et al (2008) Identification and mapping QTL for high-temperature adult-plant resistance to stripe rust in winter wheat (*Triticum aestivum* L.) cultivar ‘Stephens’. *Theor Appl Genet* 117:793–802
- Sehgal D, Yadav R (2009) Molecular markers based approaches for drought tolerance. In: *Molecular techniques in crop improvement*. Springer, Berlin, pp 207–230
- Setter TL, Yan J, Warburton M et al (2010) Genetic association mapping identifies single nucleotide polymorphisms in genes that affect abscisic acid levels in maize floral tissues during drought. *J Exp Bot* 61:303–318
- Shi L, Lv X, Weng J et al (2014) Genetic characterization and linkage disequilibrium mapping of resistance to gray leaf spot in maize (*Zea mays* L.). *The Crop J* 2:132–143
- Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: join map. *Plant J* 3:739–744
- Su Z, Hao C, Wang L et al (2011) Identification and development of a functional marker of TaGW2 associated with grain weight in bread wheat (*Triticum aestivum* L.). *Theor Appl Genet* 122:211–223
- Tan Y, Sun M, Xing Y et al (2001) Mapping quantitative trait loci for milling quality, protein content and color characteristics of rice using a recombinant inbred line population derived from an elite rice hybrid. *Theor Appl Genet* 103:1037–1045
- Tanksley SD (1993) Mapping polygenes. *Annu Rev Genet* 27:205–233
- Tanksley S, Ganal M, Prince J et al (1992) High density molecular linkage maps of the tomato and potato genomes. *Genetics* 132:1141–1160
- Tian Y, Zhang H, Xu P et al (2014) Genetic mapping of a QTL controlling leaf width and grain number in rice. *Euphytica* 1–11
- Utz H, Melchinger A (1996) PLABQTL: a program for composite interval mapping of QTL. *J Quant Trait Loci* 2:1–5
- Van Inghelandt D, Melchinger AE, Martinant J-P et al (2012) Genome-wide association mapping of flowering time and northern corn leaf blight (*Setosphaeria turcica*) resistance in a vast commercial maize germplasm set. *BMC Plant Biol* 12:56
- Vanniarajan C, Vinod K, Pereira A (2012) Molecular evaluation of genetic diversity and association studies in rice (*Oryza sativa* L.). *J Genet* 91:9–19
- Vijayalakshmi K, Fritz AK, Paulsen GM et al (2010) Modeling and mapping QTL for senescence-related traits in winter wheat under high temperature. *Mol Breed* 26:163–175
- Wang A-Y, Li Y, Zhang C-Q (2012a) QTL mapping for stay-green in maize (*Zea mays*). *Can J Plant Sci* 92:249–256
- Wang M, Yan J, Zhao J et al (2012b) Genome-wide association study (GWAS) of resistance to head smut in maize. *Plant Sci* 196:125–131
- Wen W, Mei H, Feng F et al (2009) Population structure and association mapping on chromosome 7 using a diverse panel of Chinese germplasm of rice (*Oryza sativa* L.). *Theor Appl Genet* 119:459–470
- Weng J, Xie C, Hao Z et al (2011) Genome-wide association study identifies candidate genes that affect plant height in Chinese elite maize (*Zea mays* L.) inbred lines. *PLoS ONE* 6:e29229
- Whitt SR, Buckler IV ES (2003) Using natural allelic diversity to evaluate gene function. In: *Plant functional genomics*. Springer, Berlin, pp 123–139
- Wilson LM, Whitt SR, Ibáñez AM et al (2004) Dissection of maize kernel composition and starch production by candidate gene association. *Plant Cell Online* 16:2719–2733
- Xu F-F, Tang F-F, Shao Y-F et al (2014) Genotype × environment interactions for agronomic traits of rice revealed by association mapping. *Rice Sci* 21:133–141

- Xue Y, Warburton ML, Sawkins M et al (2013) Genome-wide association analysis for nine agronomic traits in maize under well-watered and water-stressed conditions. *Theor Appl Genet* 126:2587–2596
- Yan WG, Li Y, Agrama HA et al (2009) Association mapping of stigma and spikelet characteristics in rice (*Oryza sativa* L.). *Mol Breed* 24:277–292
- Yan J, Kandianis CB, Harjes CE et al (2010) Rare genetic variation at *Zea mays* crtRB1 increases [beta]-carotene in maize grain. *Nat Genet* 42:322–327
- Yang N, Lu Y, Yang X et al (2014) Genome wide association studies using a new nonparametric model reveal the genetic architecture of 17 agronomic traits in an enlarged maize association panel. *PLoS Genet* 10:e1004573
- Yao H, Zhou Q, Li J et al (2002) Molecular characterization of meiotic recombination across the 140-kb multigenic a1-sh2 interval of maize. *Proc Natl Acad Sci* 99:6157–6162
- Yoshida T, Nishida H, Zhu J et al (2010) Vrn-D4 is a vernalization gene located on the centromeric region of chromosome 5D in hexaploid wheat. *Theor Appl Genet* 120:543–552
- Young ND (1994) Constructing a plant genetic linkage map with DNA markers. In: DNA-based markers in plants. Springer, Berlin, pp 39–57
- Young N (1996) QTL mapping and quantitative disease resistance in plants. *Annu Rev Phytopathol* 34:479–501
- Yu J, Pressoir G, Briggs WH et al (2005) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* 38:203–208
- Yu L-X, Morgounov A, Wanyera R et al (2012) Identification of Ug99 stem rust resistance loci in winter wheat germplasm using genome-wide association analysis. *Theor Appl Genet* 125:749–758
- Zegeye H, Rasheed A, Makdis F et al (2014) Genome-wide association mapping for seedling and adult plant resistance to stripe rust in synthetic hexaploid wheat. *PLoS ONE* 9:e105593
- Zeng Z-B (1994) Precision mapping of quantitative trait loci. *Genetics* 136:1457–1468
- Zhang J, Hao C, Ren Q et al (2011) Association mapping of dynamic developmental plant height in common wheat. *Planta* 234:891–902
- Zhang Y, Xu L, Fan X et al (2012) QTL mapping of resistance to gray leaf spot in maize. *Theor Appl Genet* 125:1797–1808
- Zhang X, Tang B, Yu F et al (2013) Identification of major QTL for waterlogging tolerance using genome-wide association and linkage mapping of maize seedlings. *Plant Mol Biol Rep* 31:594–606
- Zheng P, Allen WB, Roesler K et al (2008) A phenylalanine in DGAT is a key determinant of oil content and composition in maize. *Nat Genet* 40:367–372
- Zhu C, Yu J (2009) Nonmetric multidimensional scaling corrects for population structure in association mapping with different sample types. *Genetics* 182:875–888
- Zhu C, Gore M, Buckler ES et al (2008) Status and prospects of association mapping in plants. *Plant Genome* 1:5–20

Molecular Breeding for Sustainable Crop Improvement
Volume 2

Rajpal, V.R.; Rao, S.R.; Raina, S.N. (Eds.)

2016, XI, 478 p. 56 illus., 12 illus. in color., Hardcover

ISBN: 978-3-319-27088-3