



Review Article

Association Mapping: A Step Forward to Discovering New Alleles for Crop Improvement

AHMED YAHYA AL-MASKRI, MUHAMMAD SAJJAD[†] AND SULTAN HABIBULLAH KHAN^{†1}

Department of Crop Sciences, College of Agricultural and Marine Sciences, Sultan Qaboos University, Sultanate of Oman

[†]Centre of Agricultural Biochemistry and Biotechnology, University of Agriculture, Faisalabad, Pakistan

¹Corresponding author's e-mail: sultan@uaf.edu.pk

ABSTRACT

Understanding the basis of complex quantitative traits of economic importance is a major tactic behind the whole progress in plant breeding. Reading variation in phenotypes in the form of changes in DNA sequence and their location in the genome has been, traditionally, performed using linkage mapping, where crosses are made to develop a family with known parentage followed by identification of genetic markers co-segregating with target traits within this family. During the past decade, association mapping, commonly applied in human genetics, has shown greater promise and power of mapping complex quantitative traits in plants as compared to traditional linkage mapping. The focus of this review is to provide plant breeding community with the basic concept, pros and cons along with a brief description of some pioneer and recent reports to reveal the feasibility, achievements and future prospects of association mapping in plant genetics. The review would be useful for students and professional scientists working in the field of plant sciences to recognize the applications of this new approach and utilize in crop improvement programs both in academia and research. © 2012 Friends Science Publishers

Key Words: Gene mapping; Linkage disequilibrium; Association mapping; Plant breeding

INTRODUCTION

The germplasm is often exploited to develop improved crop varieties for changing needs and environments (Sajjad *et al.*, 2011). The vast amount of genetic variation present in the form of germplasm can be exploited to the best once the traits of economic importance have been mapped with molecular markers (Akbari *et al.*, 2006). Mapping a nucleotide sequence underlying a specific trait offers an opportunity for plant breeders to apply marker assisted (MAS) selection. Most of yield contributing traits are controlled by many loci and their molecular characterization and genetic mapping is called quantitative trait loci mapping (QTL-mapping). The objective of gene mapping is to find molecular markers, which are neutrally inherited and closely linked due to location within or in close proximity of the genes governing the quantitative traits. Two approaches are mostly used for genetic mapping: (i) Family-based linkage mapping (ii) LD-based Association mapping (Mackay & Powell, 2007). Family-based linkage (FBL) mapping is classical approach in which LD is created by developing a population by crossing few founders. For FBL mapping, the first step is to establish mapping populations like F_2 , double haploids, back crosses, recombinant inbred lines and near isogenic lines, which are then phenotyped to find out segregation of the trait in different environments. In the next step, DNA markers showing polymorphism between the

parents and among segregants are identified. For this, a set of markers is screened for polymorphism and the polymorphic markers are used to generate genotypic data to construct linkage map (relative genetic distance) and order (position) of the molecular markers used for genotyping. The genetic map is accomplished by the assessment of recombination frequencies between the markers. The markers located on the linkage map are associated with the phenotypic data of trait(s) being studied and significantly correlated markers with a phenotypic trait are considered to be closely linked with the QTL region affecting the trait being mapped.

In family based linkage analysis, the accuracy of mapping a gene relies on the size of mapping population, genetic variation covered by the population, and number of molecular markers applied. After, QTLs underlying a specific trait are exactly tagged with molecular markers using FBL mapping approach, the markers can be used to transfer the gene of interest from a donor line to the target genotype (marker assisted selection). Even though, linkage mapping is being used for gene mapping in crop plants, it is very costly, has low resolution and evaluates few alleles simultaneously in a relatively longer time scale (Flint-Garcia *et al.*, 2003; Gupta *et al.*, 2005; Stich *et al.*, 2006; Ross-Ibarra *et al.*, 2007). Low resolution in FBL mapping is due to lower number of meiotic events happened since experimental crossing in the near past (Jannink & Walsh,

2002). Linkage disequilibrium-based association mapping offers an alternative approach to overcome the limitations of FBL approach (Ross-Ibarra *et al.*, 2007).

Association mapping: Current gene mapping efforts are shifting from conventional FBL based mapping to LD based association mapping (Goldstein & Weale, 2001), which is the most effective approach to utilize natural variation in the form of *ex situ* conserved crop genetic resources. In association mapping, a natural population is surveyed to determine marker-trait associations using LD (Flint-Garcia *et al.*, 2003). LD refers to historically increased non-equilibrium (reduced level of recombinations) of specific alleles at various loci. The level of LD extent can be measured statistically and has been widely applied in humans to map and finally clone genes controlling complex genetic traits (Risch & Merikangas, 1996; Weiss & Clark, 2002; Chapman *et al.*, 2003; Taniguchi *et al.*, 2006). This approach was extended to plants in 2001 and a substantially increased mapping resolution over F₁-derived mapping populations was reported (Thornsberry *et al.*, 2001).

Association mapping offers several advantages over conventional FBL mapping. The availability of huge genetic variation in the form of germplasm provides broader allele coverage and saves time and cost to establish tedious and expensive biparental mapping populations, and most importantly offers higher resolution due to the exploitation of relatively higher number of meiotic events throughout the history of germplasm development. Furthermore, AM also offers the possibility of using historically measured phenotypic data (Kraakman *et al.*, 2004; Kraakman *et al.*, 2006).

The general approach of association mapping (AM) includes six steps as outlined in Fig. 1. (i) a collection of diverse genotypes are selected that may include, land races, elite cultivars, wild relatives and exotic accessions, (ii) a comprehensive and precise phenotyping is performed over the traits such as, yield, stress tolerance or quality related traits of the selected genotypes in multiple repeats and years/environments, (iii) the genotypes are then scanned with suitable molecular markers (AFLP, SSRs, SNPs), (iv) population structure and kinships are determined to avoid false positives followed by (v) quantification of LD extent using different statistics like D , D' or r^2 . Finally, (vi) genotypic and phenotyping data are correlated using appropriate statistical software allowing tagging of molecular marker positioned in close proximity of gene(s) underlying a specific trait. Consequently, the tagged gene can be mobilized between different genotypes and/or cloned and annotated for a precise biological function.

Concept of linkage disequilibrium: Coinheritance of two or more loci on a chromosome within a genetic region is called genetic linkage. In population genetics, linkage equilibrium and linkage disequilibrium describe linkage of alleles at different loci. Linkage equilibrium is the random combinations of alleles at various loci where observed haplotype frequencies agree with the predicted haplotype

frequencies in a population, which is equal to the product of individual frequency of genetic markers in each haplotype. LD, on the other hand, is a non-random association of alleles at various loci and describes non equal haplotypes frequency in a population. Mathematically, linkage equilibrium can be described as $PAB = PA \times PB$ and linkage disequilibrium as $PAB \neq PA \times PB$ (Gupta *et al.*, 2005), where A and B are alleles at two different loci, PAB is the frequency of haplotypes having both alleles at the two loci, PA and PB are the frequency of haplotypes having only A allele and B allele, respectively.

Concept of linkage disequilibrium is different from that of linkage although strong linkage is the major cause of LD between alleles. Significant LD can occur between alleles at distant loci or even at different chromosomes, generated by different genetic factors other than linkage (Huttley *et al.*, 1999; Stich *et al.*, 2006). In literature, LD is also described as “gametic disequilibrium” (GD) or “gametic phase disequilibrium (GPD), describing the same non-random association of alleles in a population of unrelated individuals, assuming Hardy-Weinberg law (Jannink & Walsh, 2002).

Linkage disequilibrium was first defined by Jennings in 1917 and quantified in 1964 by Lewtonin (Abdurakhmonov & Abdurkarimov, 2008). The generally used LD measures include D , D' , r^2 , D^2 , D^* , F and δ (Hedrick, 1987; Devlin & Risch, 1995; Jorde, 1995; Gupta *et al.*, 2005). The details of their formulae have been reviewed by Flint-Garcia *et al.* (2003), Gupta *et al.* (2005) and Oraguzie *et al.* (2007). Selection of suitable LD measure depends on the objective, situation and case of study, however, D' and r^2 are mostly used LD measures (Gupta *et al.*, 2005; Oraguzie *et al.*, 2007). The statistics D' and r^2 perform differently under different conditions, reflecting various aspect of LD (Flint-Garcia, 2003). This can be seen in Fig. 2, where two alleles are correlated resulting into absolute LD.

In this case two linked mutations occur, simultaneously, without any recombination. In this instance, mutation and recombination history is the same and D' and r^2 have same value of 1 (Fig. 2a). When the recombination and mutation histories of the haplotypes are different, then, D' and r^2 perform differently, (Fig. 2b). In case where alleles are in linkage equilibrium and the recombination history is different for haplotypes but mutation history is the same, both D' and r^2 will be zero (Fig. 2c).

Factors affecting LD: Several demographic and genetic factors play their role to shape the haplotypic LD blocks in genome (Gupta *et al.*, 2005; Stich *et al.*, 2005, 2006, 2007; Oraguzie *et al.*, 2007) out of which mutation and recombination are the key factors affecting LD significantly. Increased LD is the result of new mutations, population structure, autogamy, genetic isolation, admixture, genetic drift, small founder population size, epistasis, genomic rearrangement, selection and kinship, whereas higher rates of recombination and mutation, recurrent mutations, gene

conversion and outcrossing significantly decrease LD (Gupta *et al.*, 2005; Oraguzie *et al.*, 2007).

Successful association mapping depends on LD because of genetic linkage. Nevertheless, significant LD is often observed between distant loci located even on different chromosomes causing spurious associations in AM studies. The long stretched or genetically unlinked LD is the result of LD causing factors other than genetic linkage (Stich *et al.*, 2005 & 2006). The factors that cause LD between unlinked loci through “hitchhiking” effect include population structure, population admixture, selection, epistatic or co-selection of co-adapted genes (Cannon, 1963) during breeding for multiple traits (Wang *et al.*, 2002; Oraguzie *et al.*, 2007). Theoretically, kinship creates LD between genetically linked loci but it can also create LD between genetically unlinked loci when predominant parents are included in the population. In maize, kinship equally generated LD between genetically linked and unlinked loci (Stich *et al.*, 2005). The factors including genetic drift and bottlenecks also generate LD (Stich *et al.*, 2005; Stich *et al.*, 2007) resulting non-uniform distribution of LD in genome (Huttley *et al.*, 1999).

Knowing the factors affecting LD, an obvious question is whether increased or decreased level of LD favors AM? The extensive level of LD (long stretched LD) reduces the number of markers required for marker-trait association but lowers the mapping resolution (coarse mapping). Conversely, less extensive LD (short stretched LD) needs relatively more number of markers to mine a gene but increase mapping resolution. Selection of a population with LD level higher or lower depends on the objective of mapping study.

The “ascertainment biasness” is another reason that affects the LD. Some of these factors are discussed in detail by Gupta *et al.* (2005). Minor alleles (present in less than 10% individuals) largely inflate LD values (Caldwell *et al.*, 2006). Hence, in LD quantification and AM, markers with minor allele frequency are (i) replaced with missing values (Barnaud *et al.*, 2006; Brescghello & Sorrels, 2006) (ii) pooled into common allele class (Hamblin *et al.*, 2004) or (iii) removed before analysis (Hamblin *et al.*, 2004; Kraakman *et al.*, 2004; Caldwell *et al.*, 2006; Kraakman *et al.*, 2006).

Quantification of LD in plants: Quantification of LD have been an important research objective in model plant *Arabidopsis* and other major crop plants including barley, rice, durum wheat, maize, spring wheat, sorghum, soybean and sugarcane. Nordborg *et al.* (2002) quantified LD around the flowering time gene in a set of 20 *Arabidopsis* genotypes and found LD decay within a distance of 1cM (250 kb). Later on, Nordborg *et al.* (2005) genotyped a global collection of 96 *Arabidopsis* genotypes in which LD extended up to 50-250 kb. In Michigan genotypes, LD ranged from 50 cM to 100 cM. This long stretched LD was explained as founder effect or genetic bottle neck. In contrast, Tian *et al.* (2002) targeted the region of *rps5* gene

and reported low LD extent up to only 10 kb. Likewise, Shepard and Purugganan (2003) found that LD decayed within 10-50 kb distance in the *CLAVATA2* segment of *Arabidopsis*. Ehrenreich *et al.* (2007) extensively sequenced 600 bp segments of *MAX2* and *MAX3* regions in a panel of 96 *Arabidopsis* genotypes collected from countries of Central Europe and reported LD decay within ~10kb. Plagnol *et al.* (2006) conducted a whole-genome scan of 1347 fragments of 600 bp and determined the complete disappearance of LD after ~100 kb.

A rapid LD decay was reported in maize, a highly cross pollinated species. The first LD extent in maize was determined by Tenaillon *et al.* (2001). They genotyped 21 loci over the 25 exotic and United States genotypes and found LD decay within 400 bp with $r^2 = 0.2$. The extent of LD was 1kb in US genotypes. Later, Remington *et al.* (2001) surveyed 6 genes in a panel of 102 tropical and subtropical inbred lines and reported very rapid LD decay within 200-2000 bp ($r^2 = 0.1$) with the exception *su1* gene where LD was significant ($r^2 = 0.3-0.4$) for a length of 10 kb. The significant LD in *su1* loci was explained by string selection in *su1* gene. The same set of genotypes was genotyped with 47 SSRs (Remington *et al.*, 2001) and higher extent of LD was found than those found with SNP data. The difference of LD extent was because of dissimilar mutation rate of the two marker types. Long stretches of LD were measured by Thornsberry *et al.* (2001) in and around *Dwarf8* gene. They found long stretches of LD (up to 3 kb) were interspersed with regions of relatively low LD. Jung *et al.* (2004) surveyed *adh1* locus and reported LD extent within 500 kb. A set of 147 European and US inbred lines was examined with 100 SSRs by Stich *et al.* (2005). They reported significant LD blocks of 41 cM in dent group and 26 cM for flint group. The LD distribution among all chromosomes was non-uniform. Stich *et al.* (2006) reported relatively much smaller average LD level (4 cM) for AFLP but higher LD extent (30-31 cM) with SSRs in a panel of 72 European inbred lines consisting of both flint and dent types. This result emphasizes the use of both AFLP and SSR marker systems in association mapping with better results from SSRs because of better LD detection.

Whole-genome LD has also been measured in various crop species extending up to 10-50 cM in soybean (Zhu *et al.*, 2003; Jun *et al.*, 2007), 10 cM in sugarcane (Flint-Garcia *et al.*, 2003), 3cM in sugarbeet (Kraft *et al.*, 2000), 5-10 cM in grape (Barnaud *et al.*, 2006), 50 cM in sorghum (Hamblin *et al.*, 2004), 16-34kb in poplar (Yin *et al.*, 2004), 200 bp in silage maize (Guillet-Claude *et al.*, 2004) and 0.5 kb - 2 kb in rye grass (Skot *et al.*, 2005; Ponting *et al.*, 2007; Xing *et al.*, 2007). The LD extent in cotton was studied by Abdurakhmonov *et al.* (2007) who surveyed 200 SSRs in germplasm of 335 genotypes and reported whole-genome extended up to 25 cM with $r^2 > 0.1$.

In self-pollinated plants, higher levels of LD were observed relative to outcrossing crop species (Gupta *et al.*, 2005; Abdurakhmonov & Abdurkarimov, 2008).

Table I: Some successful reports of Association mapping in plants

Species	Mapped traits	References
Self-Pollinated		
Arabidopsis	Growth response, flowering time, branching architecture and pathogen resistance	Caicedo <i>et al.</i> (2004); Oslen <i>et al.</i> (2004); Ehrenreich <i>et al.</i> (2007); Zhao <i>et al.</i> (2007)
Rice	Plant height, flag leaf length and width, tiller number, stem diameter, stigma characteristics, flowering date, panicle length, grain length and width, grain thickness, 1000-grain weight.	Zhang <i>et al.</i> (2005); Agrama <i>et al.</i> (2007); Yan <i>et al.</i> (2009)
Barley	Plant height, heading date, flowering date, rachilla length, yield stability, yield, mildew and leaf rust resistance.	Igartua <i>et al.</i> (1999); Ivandic <i>et al.</i> (2003); Kraakman <i>et al.</i> (2004); Kraakman <i>et al.</i> (2006)
Wheat	Plant height, milling quality, High molecular weight glutenin, 1000-kernal weight, protein contents, drought tolerance, sedimentation value, test weight, starch concentration, insect and disease resistance.	Breseghello and Sorrells (2006); Ravel <i>et al.</i> (2005); Roy <i>et al.</i> (2006); Crossa <i>et al.</i> (2007); Peng <i>et al.</i> (2009); Reif <i>et al.</i> (2011); Adhikari <i>et al.</i> (2011); Zhang <i>et al.</i> (2011).
Potato	Resistance to wilt disease, phytophthora, bacterial blight, tuber shape, flesh color, under water weight and maturity etc.	Gebhardt <i>et al.</i> (2004); Simko <i>et al.</i> (2004); Malosetti <i>et al.</i> (2007); D'hoop <i>et al.</i> (2008)
Soybean	Seed protein contents	Jun <i>et al.</i> (2007)
Cross pollinated		
Maize	Plant height, endosperm color, starch production, flowering time, maysin and chlorogenic accumulation forage quality, cell wall digestibility and oleic acid concentration.	Remington <i>et al.</i> (2001); Thornsberry <i>et al.</i> (2001); Guillet-Glaude <i>et al.</i> (2004); Wilson <i>et al.</i> (2004); Anderson <i>et al.</i> (2007)
Forage grasses	Cold tolerance, forage quality, flowering time and carbohydrate contents	Dobrowolski and Forster (2007); Skot <i>et al.</i> (2007)
Forest trees	Early-wood micro fibril angle, wood growth rate, and wood density.	Thumma <i>et al.</i> (2005); Wicox <i>et al.</i> (2007)

In durum wheat (*Triticum turgidum*) Maccaferri *et al.* (2005) examined LD in 134 genotypes that extended up to 10-20cM with $D' = 0.67$ and 0.43 , respectively.

In rice (*Oryza sativa* L.), Garris *et al.* (2003) determined LD within 100kb with $r^2 = 0.1$ around locus *Xa5* using 21 SSRs in a set of 114 lines. A global assembly of rice (*O. sativa* L.) studied with 176 SSRs showed long range LD decay within ~50cM ($D' = 0.5$) in the US and global pool (Agrama & Eizenga, 2008). In overall global collections of land races and wild relatives, LD persisted up to 225 cM with significant $D' > 0.5$. Conversely, some studies described relatively smaller extent of LD in genetically broad based germplasm and high level of LD in narrow based germplasm (Remington *et al.*, 2001; Stich *et al.*, 2006). The LD decay in different rice species varied significantly. The LD decay in *O. rufipogon* was at 5 kb and in *O. sativa* it was at 50 kb (Rakshit *et al.*, 2007). Mather *et al.* (2007) reported greatest LD in temperate japonica rice (> 500 kb) followed by tropical japonica rice (175 kb) and lowest in indica type (75 kb) using SNPs. Recently, Yan *et al.* (2009) assayed 90 accessions developed from USDA and evaluated genome wide LD through pairwise comparisons of 109 loci. Out of 4,046 estimates by r^2 , 312 (7%) were significant ($r^2 > 0.2$).

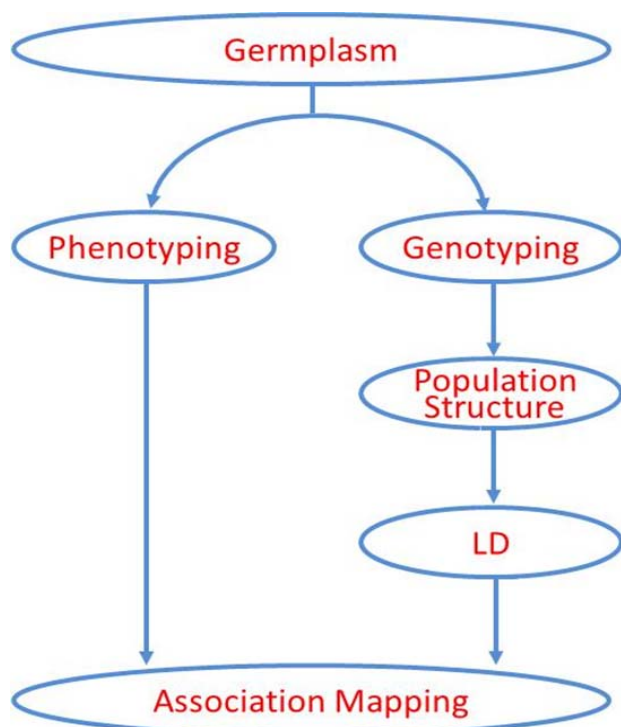
The LD extent in barley ranged from 10 cM to 50 cM depending on marker system and the genotypes (Kraakman *et al.*, 2004; Malysheva-Otto, 2006). Caldwell *et al.* (2006) determined LD around hardness locus in three populations of barley and observed LD extent up to 212 kb in elite inbred lines and 98 kb in land races. Morrell *et al.* (2005) reported a rapid decline of LD within 300 bp around 18 nuclear genes in 25 diverse wild accessions.

In hexaploid wheat (*Triticum aestivum* L.), Berseghello and Sorrells (2006) surveyed a panel of 95 winter wheat accessions with 36 SSR markers. They

observed strong LD within < 5 cM and ~ 1 cM for centromeric region of chromosome 5A and chromosome 2D, respectively. Tommasini *et al.* (2007) surveyed 91 SSRs and STS markers in 44 varieties and 240 RILs. They observed LD on chromosome 3B stretched up to 0.5 cM in 44 cultivars and 30 cM in 240 RILs. This study favors the usefulness of germplasm over bi-parental populations in association mapping. Crossa *et al.* (2007) analyzed LD in a panel of 170 wheat lines derived from five CIMMYT elite spring wheat yield trials (ESWYTs) using 813 DArT markers and 831 other markers including SSRs, AFLPs and RFLs. Across all chromosomes a total of 43 LD blocks were observed with an average length of 9.93 cM. The longest LD blocks were observed on chromosomes 1B and 4B with approximate length of 64.96 and 87.41 cM, respectively. Conversely, very low LD extent of 2.3 cM was observed on chromosome 2A in a set of 137 Chinese wheat germplasm accessions assayed with 37 SSRs covering 2A with 3.75 average marker intervals (Yao *et al.*, 2009). The highest extend of LD, as determined by 394 SNPs in population of 478 wheat genotypes was found in D-genome (20.8-27.1 cM), followed by smaller LD in B-genome (20.4-21.5) and A-genome (16.7-19.8 cM) (Chao *et al.*, 2010). These reports revealed that LD structure varies with populations and type of marker.

Implication of LD quantification for AM: The important implications from the above reports on LD quantification in plants are: (i) LD declines quickly in outcrossing species facilitating fine mapping of a trait of interest. Furthermore, LD declines rapidly in a group of unrelated individuals (germplasm) even in self-pollinated plant species as compared to mapping populations resulting from bi-parental crosses, that offers benefit of discovery of new alleles (Tommasini *et al.*, 2007); (ii) the LD extent varies across genomic regions and populations (iii) LD extent varying

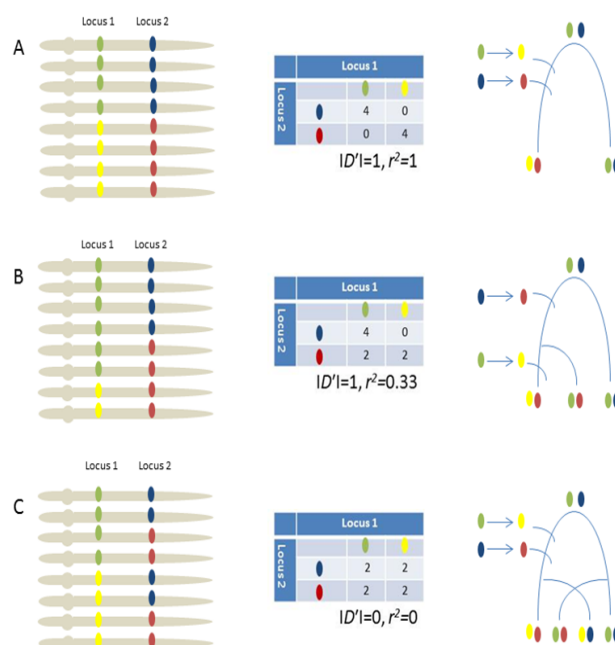
Fig. 1: A simplified flowchart showing different stages of association mapping for tagging a gene of interest using germplasm accessions



with marker systems capturing different historic changes in a genome because of varying mutation rates (iv) an average estimate of whole-genome LD extent in gene pool may not represent LD extent of specific genomic region or specific population. Therefore, the specific genomic regions or populations should be additionally explored for marker-trait associations in that specific genomic region or population group (v) LD extent in narrow-based germplasm is longer than LD extent in broad-based germplasm. This implies that for higher resolution the mapping population should be composed of unrelated diverse genotypes (Remington *et al.*, 2001; Stich *et al.*, 2006). This LD structure in narrow and broad-based germplasm also provides a strategy for coarse mapping in narrow-based germplasm with less number of markers and then fine mapping employing higher number of markers in germplasm with broader genetic base; and (vi) biological behavior and confounding population behavior seriously effect the patterns of LD in plant germplasm and hence must be taken into account for valid association studies.

Power of association mapping: In a set of unrelated individuals, mapping power using association mapping approach is the probability of detecting the true marker-trait associations that depends on (i) the evolution and extent of LD in the genomic region harboring the loci for trait(s) being mapped and mapping population; (ii) the type of gene action of the trait; (iii) size and composition of population; (iv) field design and accuracy of phenotyping, genotyping and data

Fig. 2: Hypothetical description of linkage disequilibrium (LD) between linked loci caused by different mutations and recombinations showing the behavior of the D' and r^2 statistics. The left column represents the allelic states of two genes. The middle column shows the 2x2 square table of haplotypes and the resulting D' and r^2 statistics. The right column depicts a putative tree responsible for the observed LD. (A) Absolute LD prevails when two loci have a similar mutational history without recombination. Both D' and r^2 equals 1. (B) LD can exist when mutations occur on different lineages with no recombination between two loci. Notice the difference in LD measures as calculated by D' and r^2 . (C) LD is produced when there is recombination between loci, regardless of mutational history. In this situation, both D' and r^2 equal 0



analysis. The power of AM can be increased by better data recording and analysis and increasing population size. In AM there are specific statistical methods to determine the false-positives (Type 1 error) such as permutation (Churchill & Doerge, 1994) or false recovery rates (Benjamini & Hochberg, 1995). For AM study in the presence of population structure Pritchard *et al.* (2000) established a useful technique for structured association (SA). Structured association (SA) uses Bayesian approach (Marttinen & Corander, 2010) to search sub-populations using Q matrix to avoid false positives. Population structure (Q-matrix) and kinship coefficient (K-matrix) can be estimated in subpopulations using the program STRUCTURE (Pritchard & Wen, 2004). Recently, Yu *et al.* (2006) established another approach called a mixed linear model (MLM) to bloc structure information (Q-matrix) and kinship information (K-matrix) in AM analysis. Later on, the Q+K MLM model performed better even in highly structured

population of *Arabidopsis* as compared to any other model that used Q- or K-matrix alone (Zhao *et al.*, 2007).

Some mixed model approaches also combine QTL and LD, where, QTLs or already known genes are used as a priori information in association mapping (Thumma *et al.*, 2005). This is the effective approach in AM that reduces the number of markers and populations size. This approach also increases the precision and power of marker-trait associations (Ball, 2005).

Selected reports on AM in plants: First reports of AM in plants emerged in 1996 in rice and 1997 in oat by Virk *et al.* (1996) and Beer *et al.* (1997) respectively. Virk *et al.* (1996) linked six traits with RAPD markers whereas Beer *et al.* (1997) applied RFLPs to map 13 quantitative traits without taking population structure into account that resulted in more spurious associations. However, the first report of AM in plants accounting for population structure was published on maize in 2001 (Thornsberry *et al.*, 2001). Later on, AM was extended to barley, *Arabidopsis*, potato, rice, wheat and sea beet considering population structure and extent of LD level. In barley traits of agronomic importance including grain yield, yield stability, flowering time, heading time, plant stature, and resistance to diseases were associated with various kinds of molecular markers (Igaruta *et al.*, 1999; Ivandic *et al.*, 2003; Kraakman *et al.*, 2004; Kraakman *et al.*, 2006). In hexaploid wheat first study on AM was reported by Breseghello and Sorrells (2006) tagging SSRs with kernel size and milling quality. To date AM also has been extended to long life-span forest plants (Wilcox *et al.*, 2007). There are also successful reports on AM for different agronomic and physiological traits in *Arabidopsis*, rice, soyabean, maize and forage crops that are summarized in Table I.

CONCLUSION

Association mapping after its successful application in human genetics has found its way in plant genetics to help decipher complex quantitative traits. Addition of LD-based powerful association mapping tools has demonstrated a remarkable flourish in the field of crop genomics. With the recent methodological developments to minimize spurious associations in structured populations, applications of AM has been extended from model plant *Arabidopsis* to field crops such as rice, wheat, maize, barley, sugarcane and forage grasses. The increasing number of AM studies in crop species indicates the potential of this approach in all plant species in near future. Furthermore, advancements to develop more cost-effective sequencing technologies for efficient genome sequencing of crop plants will certainly accelerate progress in genome-wide association studies (Anonymous, 2007) discovering rare and common alleles (Estivill & Armengol, 2007) and epigenomic information about the trait of interest. This will increase the power of LD-based association mapping for discovering true associations to facilitate its effective utilization in crop breeding programs.

REFERENCES

- Abdurakhmonov, A.Y. and A. Abdurkarimov, 2008. Application of association mapping to understanding the genetic diversity of plant germplasm resources. *Int. J. Plant Genom.*, 2008: 1–18
- Abdurakhmonov, I.Y., R.J. Kohel, S. Saha, A.E. Pepper, J.Z. Yu, Z.T. Buriev, A. Abdullaev, S. Shermatov, J.N. Jenkins, B. Scheffler and A. Abdurkarimov, 2007. Genomewide linkage disequilibrium revealed by microsatellite markers and association study of fiber quality traits in cotton. *Proceedings of the 15th Plant and Animal Genome Conference*, San Diego, California, USA
- Adhikari, T., E.W. Jackson, S. Gurung, J. Hansen and J.M. Bonman, 2011. Association mapping of quantitative resistance to phaeosphaeria nodorum in spring wheat landraces from the USDA national small grains collection. *Phytopathology*, 101: 1301–1310
- Agrama, H.A. and G.C. Eizenga, 2008. Molecular diversity and genome-wide linkage disequilibrium patterns in a worldwide collection of *Oryza sativa* and its wild relatives. *Euphytica*, 60: 339–355
- Akbari, M., P. Wenzl, V. Caig, J. Carling, L. Xia, S. Yang, G. Uszynski, V. Mohler, A. Lehmsiek, H. Kuchel, M.J. Hayden, N. Howes, P. Sharp, P. Vaughan and B. Rathmell, 2006. Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. *Theor. Appl. Genet.*, 113: 1409–1420
- Anderson, J.R., I. Zein, G. Wenzel, B. Krützfeldt, J. Eder, M. Ouzunova and T. Lübberstedt, 2007. High levels of linkage disequilibrium and associations with forage quality at a Phenylalanine Ammonia-Lyase locus in European maize (*Zea mays* L.) inbreds. *Theor. Appl. Genet.*, 114: 307–319
- Anonymous, 2007. The Wellcome Trust Case Control Consortium, “Genomewide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447: 661–678
- Ball, R.D., 2005. Experimental designs for reliable detection of linkage disequilibrium in unstructured random population association studies. *Genetics*, 170: 859–873
- Barnaud, A., T. Lacombe and A. Doligez, 2006. Linkage disequilibrium in cultivated grapevine, *Vitis vinifera* L. *Theor. Appl. Genet.*, 112: 708–716
- Beer, S.C., W. Siripoonwivat, L. S. O’donoghue, E. Souza, D. Matthews and M.E. Sorrells, 1997. Associations between molecular markers and quantitative traits in an oat germplasm pool: can we infer linkages? *J. Agric. Genom.*, 3: 197
- Benjamini, Y. and Y. Hochberg, 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Royal Stat. Soc. Series B.*, 57: 289–300
- Breseghello, F. and M.E. Sorrells, 2006. Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics*, 172: 1165–1177
- Caicedo, A., J.R. Stinchcombe, K.M. Olsen, J. Schmitt and M.D. Purugganan, 2004. Epistatic interaction between *Arabidopsis* *FRI* and *FLC* flowering time genes generates a latitudinal cline in a life history trait. *Proc. Natl. Acad. Sci.*, 101: 15670–15675
- Caldwell, K.S., J. Russell, P. Langridge and W. Powell, 2006. Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. *Genetics*, 172: 557–567
- Cannon, G.B., 1963. The effects of natural selection on linkage disequilibrium and relative fitness in experimental populations of *Drosophila melanogaster*. *Genetics*, 48: 1201–1216
- Chao, S., J. Dubcovsky, J. Dvorak, M.C. Luo, S.P. Baezinger, R. Matnyazov, D.R. Clark, L.E. Talbert, J.A. Anderson, S. Dreisigacker, K. Glover, J. Chen, K. Kambell, P.L. Bruckner, J.C. Rudd, S. Haley, B.F. Carver, S. Perry, M.E. Sorrells and E.D. Akhunov, 2010. Population- and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (*Triticum aestivum* L.). *BMC Genomics*, 11: 727
- Chapman, J.M., J.D. Cooper, J.A. Todd and D.G. Clayton, 2003. Detecting disease associations due to linkage disequilibrium using haplotype tags: a class of tests and the determinants of statistical power. *Human Heredity*, 56: 18–31
- Churchill, G.A. and R.W. Doerge, 1994. Empirical threshold values for quantitative trait mapping. *Genetics*, 138: 963–971

- Crossa, J., J. Burgueño, S. Dreisigacker, M. Vargas, S.A. Herrera-Foessel, M. Lillemo, R.P. Singh, R. Trethowan, M. Warburton, J. Franco, M. Reynolds, J.H. Crouch and R. Ortiz, 2007. Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. *Genetics*, 177: 1889–1913
- D'hoop, B.B., M.J. Paulo, R.A. Mank, H.J. Van Eck and F. A. Van Eeuwijk, 2008. Association mapping of quality traits in potato (*Solanum tuberosum* L.). *Euphytica*, 161: 47–60
- Devlin, B. and N. Risch, 1995. A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics*, 29: 311–322
- Dobrowolski, M.P. and J.W. Forster, 2007. Linkage disequilibrium-based association mapping in forage species". In: Oraguzie, N.C., E.H.A. Rikkerink, S.E. Gardiner and H.N. De Silva (eds.), *Association Mapping in Plants*, pp: 197–210, Springer, New York, USA
- Ehrenreich, I.M., P.A. Stafford and M.D. Purugganan, 2007. The genetic architecture of shoot branching in *Arabidopsis thaliana*: a comparative assessment of candidate gene associations vs. quantitative trait locus mapping. *Genetics*, 176: 1223–1236
- Estivill, X. and L. Armengol, 2007. Copy number variants and common disorders: filling the gaps and exploring complexity in genome-wide association studies. *PLoS Genetics*, 3: 1787–1799
- Flint-Garcia, S.A., J.M. Thornsberry and E.S. Buckler, 2003. Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Biol.*, 54: 357–374
- Garris, A., S.R. McCouch and S. Kresovich, 2003. Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the xa5 locus of rice (*Oryza sativa* L.). *Genetics*, 165: 759–769
- Gebhardt, C., A. Ballvora, B. Walkemeier, P. Oberhagemann and K. Schuler, 2004. Assessing genetic potential in germplasm collections of crop plants by marker-trait association: a case study for potatoes with quantitative variation of resistance to late blight and maturity type. *Molec. Breed.*, 13: 93–102
- Goldstein, D.B. and M.E. Weale, 2001. Population genomics: linkage disequilibrium holds the key. *Curr. Biol.*, 11: 576–579
- Guillet-Claude, C., C. Birolleau-Touchard, D. Manicacci, M. Fourmann, S. Barraud, V. Carret, J.P. Martinant and Y. Barrière, 2004. Genetic diversity associated with variation in silage corn digestibility for three O-methyltransferase genes involved in lignin biosynthesis. *Theor. Appl. Genet.*, 110: 126–135
- Gupta, P.K., S. Rustgi and P.L. Kulwal, 2005. Linkage disequilibrium and association studies in higher plants: present status and future prospects. *Plant Molec. Biol.*, 57: 461–485
- Hamblin, M.T., S.E. Mitchell, G.M. White, J. Gallego, R. Kukatla, R.A. Wing, A.H. Paterson and S. Kresovich, 2004. Comparative population genetics of the panicoid grasses: sequence polymorphism, linkage disequilibrium and selection in a diverse sample of *Sorghum bicolor*. *Genetics*, 167: 471–483
- Hedrick, P.W., 1987. Gametic disequilibrium measures: proceed with caution. *Genetics*, 117: 331–341
- Huttley, G.A., M.W. Smith, M. Carrington and S.J. O'Brien, 1999. A scan for linkage disequilibrium across the human genome. *Genetics*, 152: 1711–1722
- Igartua, E., A.M. Casas, F. Ciudad, J.L. Montoya and I. Romagosa, 1999. RFLP markers associated with major genes controlling heading date evaluated in a barley germ plasm pool. *Heredity*, 83: 551–559
- Ivancic, V., W.T.B. Thomas, E. Nevo, Z. Zhang and B.P. Forster, 2003. "Associations of simple sequence repeats with quantitative trait variation including biotic and abiotic stress tolerance in *Hordeum spontaneum*. *Plant Breed.*, 122: 300–304
- Jannink, J.L. and B. Walsh, 2002. Association mapping in plant populations". In: Kang, M.S. (ed.), *Quantitative Genetics, Genomics and Plant Breeding*, pp: 59–68. CAB International, Oxford, UK
- Jorde, B., 1995. Linkage disequilibrium as a gene-mapping tool. *American J. Human Genet.*, 56: 11–14
- Jun, T.H., K. Van, M.Y. Kim, H.S. Lee and D.R. Walker, 2007. Association analysis using SSR markers to find QTL for seed protein content in soybean. *Euphytica*, 162: 179–191
- Jung, M., A. Ching, D. Bhattaramakki, M. Dolan, S. Tingey, M. Morgante and A. Rafalski, 2004. Linkage disequilibrium and sequence diversity in a 500-kbp region around the *adh1* locus in elite maize germplasm. *Theor. Appl. Genet.*, 109: 681–689
- Kraakman, A.T.W., F. Martínez, B. Mussiraliyev, F.A. Van Eeuwijk and R.E. Niks, 2006. Linkage disequilibrium mapping of morphological, resistance and other agronomically relevant traits in modern spring barley cultivars. *Mol. Breed.*, 17: 41–58
- Kraakman, A.T.W., R.E. Niks, P.M.M.M. Van Den Berg, P. Stam and F.A. Van Eeuwijk, 2004. Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. *Genetics*, 168: 435–446
- Kraft, T., M. Hansen and N.O. Nilsson, 2000. Linkage disequilibrium and fingerprinting in sugar beet. *Theor. Appl. Genet.*, 101: 323–326
- Nordborg, M., J.O. Borevitz, J. Bergelson, C.C. Berry, J. Chory, J. Hagenblad, M. Kreitman, J.N. Maloof, T. Noyes, P.J. Oefner, E.A. Stahl and D. Weigel, 2002. The extent of linkage disequilibrium in *Arabidopsis thaliana*. *Nat. Genet.* 30: 190–193
- Nordborg, M., T.T. Hu, Y. Ishino, J. Jhaveri, C. Toomajian, H. Zheng, E. Bakker, P. Calabrese, J. Gladstone, R. Goyal, M. Jakobsson, S. Kim, Y. Morozov, B. Padhukasahasram, V. Plagnol, N. Rosenberg, A.C. Shah, J.D. Wall, J. Wang, K. Zhao, T. Kalbfleisch, V. Schulz, M. Kreitman and J. Bergelson, 2005. The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol.*, 3: 1289–1299
- Mackay, I. and W. Powell, 2007. Methods for linkage disequilibrium mapping in crops. *Trends Plant Sci.*, 12: 57–63
- Malosetti, M., C.G. Van Der Linden, B. Vosman and F.A. Van Eeuwijk, 2007. A mixed-model approach to association mapping using pedigree information with an illustration of resistance to *Phytophthora infestans* in potato. *Genetics*, 175: 879–889
- Malysheva-Otto, L.V., W. Ganal and M.S. Roder, 2006. Analysis of molecular diversity, population structure and linkage disequilibrium in a worldwide survey of cultivated barley germplasm (*Hordeum vulgare* L.). *BMC Genet.*, 7: 6
- Mather, K.A., A.L. Caicedo, N.R. Polato, K.M. Olsen, S. McCouch and M.D. Purugganan, 2007. The extent of linkage disequilibrium in rice (*Oryza sativa* L.). *Genetics*, 177: 2223–2232
- Martinen, P. and J. Corander, 2010. Efficient Bayesian approach for multilocus association mapping including gene-gene interactions. *BMC Bioinformatics*, 11: 443
- Morrell, P.L., D.M. Tolen, K.E. Lundy and M.T. Clegg, 2005. Low levels of linkage disequilibrium in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite high rates of self-fertilization. *Proc. Nat. Acad. Sci. USA*, 102: 2442–2447
- Nordborg, M., J.O. Borevitz, J. Bergelson, C.C. Berry, J. Chory, J. Hagenblad, M. Kreitman, J.N. Maloof, T. Noyes, P.J. Oefner, E.A. Stahl and D. Weigel, 2002. The extent of linkage disequilibrium in *Arabidopsis thaliana*. *Nat. Genet.*, 30: 190–193
- Olsen, K.M., S.S. Halldorsdottir, J.R. Stinchcombe, C. Weinig, J. Schmitt and M.D. Purugganan, 2004. Linkage disequilibrium mapping of *Arabidopsis* CRY2 flowering time alleles. *Genetics*, 167: 1361–1369
- Oraguzie, N.C., P.L. Wilcox, E.H.A. Rikkerink and H.N. De Silva, 2007. Linkage disequilibrium". In: Oraguzie, N.C., E.H.A. Rikkerink, S.E. Gardiner and H.N. De Silva (eds.), *Association Mapping in Plants*, pp: 11–39. Springer, New York, USA
- Peng, J., H. Wang, S.D. Haley, F.B. Peairs and N.L.V. Lupton, 2007. Molecular mapping of the Russian wheat aphid resistance gene *Dn2414* in wheat. *Crop Sci.*, 47: 2418–2429
- Plagnol, V., B. Padhukasahasram, J.D. Wall, P. Marjoram, and M. Nordborg, 2006. Relative influences of crossing over and gene conversion on the pattern of linkage disequilibrium in *Arabidopsis thaliana*. *Genetics*, 172: 2441–2448
- Ponting, R.C., M.C. Drayton, N.O.I. Cogan, M.P. Dobrowolski, G.C. Spangenberg, K.F. Smith and J.W. Forster, 2007. SNP discovery, validation, haplotype structure and linkage disequilibrium in full-length herbage nutritive quality genes of perennial ryegrass (*Lolium perenne* L.). *Molec. Genet. Genom.*, 278: 585–597
- Pritchard, K.J. and W. Wen, 2004. *Documentation for Structure Software*. The University of Chicago Press, Chicago, Ill, USA
- Rakshit, S., A. Rakshit, H. Matsumura, Y. Takahashi, Y. Hasegawa, A. Ito, T. Ishii, N.T. Miyashita and R. Terauchi, 2007. Large-scale DNA polymorphism study of *Oryza sativa* and *O. rufipogon* reveals the origin and divergence of Asian rice. *Theor. Appl. Genet.*, 114: 731–743

- Ravel, C., S. Praud, A. Murigneux, L. Linossier, M. Dardevet, F. Balfourier, P. Dufour, D. Brunel and G. Charmet, 2005. Identification of Glu-B1-1 as a candidate gene for the quantity of highmolecular-weight glutenin in bread wheat (*Triticum aestivum* L.) by means of an association study. *Theor. Appl. Genet.*, 112: 738–743
- Reif, J.C., M. Gowda, H.P. Maurer, C.F.H. Longin, V. Korzun, E. Ebmeyer, R. Bothe, C. Pietsch and T. Wurschum, 2011. Association mapping for quality traits in soft winter wheat. *Theor. Appl. Genet.*, 122: 961–970
- Remington, D.L., J.M. Thornsberry, Y. Matsuoka, L.M. Wilson, S.R. Whitt, J. Doebley, S. Kresovich, M.M. Goodman and E.S. Buckler, 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc. Natl. Acad. Sci. USA*, 98: 11479–11484
- Risch, N. and K. Merikangas, 1996. The future of genetic studies of complex human diseases. *Science*, 273: 1516–1517
- Ross-Ibarra, J., P.L. Morrell and B.S. Gaut, 2007. Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proc. National Acad. Sci. USA*, 104: 8641–8648
- Roy, J.K., R. Bandopadhyay, S. Rustgi, H.S. Balyan and P.K. Gupta, 2006. Association analysis of agronomically important traits using SSR, SAMPL, and AFLP markers in bread wheat. *Curr. Sci.*, 90: 683–689
- Sajjad M., S.H. Khan and A.S. Khan, 2011. Exploitation of germplasm for grain yield improvement in spring wheat (*Triticum aestivum* L.). *Int. J. Agric. Biol.*, 13: 695–700
- Shepard, K.A. and M.D. Purugganan, 2003. Molecular population genetics of the Arabidopsis CLAVATA2 region: the genomic scale of variation and selection in a selfing species. *Genetics*, 163: 1083–1095
- Simko, I., S. Costanzo, K.G. Haynes, B.J. Christ and R.W. Jones, 2004. Linkage disequilibrium mapping of a Verticillium dahliae resistance quantitative trait locus in tetraploid potato (*Solanum tuberosum*) through a candidate gene approach. *Theor. Appl. Genet.*, 108: 217–224
- Sköt, L., J. Humphreys, M.O. Humphreys, D. Thorogood, J. Gallagher, R. Sanderson, I.P. Armstead and I.D. Thomas, 2007. Association of candidate genes with flowering time and water-soluble carbohydrate content in *Lolium perenne* (L.). *Genetics*, 177: 535–547
- Sköt, L., M.O. Humphreys, I. Armstead, S. Heywood, K.P. Sköt, R. Sanderson, I.D. Thomas, K.H. Churlton and N. Ruairaidh Sackville Hamilton, 2005. An association mapping approach to identify flowering time genes in natural populations of *Lolium perenne* (L.). *Molec. Breed.*, 15: 233–245
- Stich, B., A.E. Melchinger, H.P. Piepho, S. Hamrit, W. Schipprack, H.P. Maurer and J.C. Reif, 2007. Potential causes of linkage disequilibrium in a European maize breeding program investigated with computer simulations. *Theor. Appl. Genet.*, 115: 529–536
- Stich, B., H.P. Maurer, A.E. Melchinger, M. Frisch, M. Heckenberger, J.R. Van Der Voort, J. Peleman, A.P. Sørensen and J.C. Reif, 2006. Comparison of linkage disequilibrium in elite European maize inbred lines using AFLP and SSR markers. *Molec. Breed.*, 17: 217–226
- Stich, B., A.E. Melchinger, M. Frisch, H.P. Maurer, M. Heckenberger and J.C. Reif, 2005. Linkage disequilibrium in European elite maize germplasm investigated with SSRs. *Theor. Appl. Genet.*, 111: 723–730
- Taniguchi, H., C.E. Lowe, J.D. Cooper, D.J. Smyth, R. Bailey, S. Nutland, B.C. Healy, A.C. Lam, O. Burren, N.M. Walke, L.J. Smink, L.S. Wicker and J.A. Todd, 2006. Discovery, linkage disequilibrium and association analyses of polymorphisms of the immune complement inhibitor, decayaccelerating factor gene (DAF/CD55) in type 1 diabetes. *BMC Genet.*, 7: 22
- Tenaillon, M.I., M.C. Sawkins, A.D. Long, R.L. Gaut, J.F.D. Doebley and B.S. Gaut, 2001. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays ssp. mays* L.). *Proc. Natl. Acad. Sci. USA*, 98: 9161–9166
- Thornsberry, J.M., M.M. Goodman, J. Doebley, S. Kresovich, D. Nielsen and E.S. Buckler, 2001. *Dwarf8* polymorphisms associate with variation in flowering time. *Nat. Genet.*, 28: 286–289
- Thumma, B.R., M.F. Nolan, R. Evans and G.F. Moran, 2005. Polymorphisms in cinnamoyl CoA reductase (CCR) are associated with variation in microfibril angle in *Eucalyptus* spp. *Genetics*, 71: 1257–1265
- Tian, D., H. Araki, E. Stahl, J. Bergelson and M. Kreitman, 2002. “Signature of balancing selection in Arabidopsis”, *Proc. Natl. Acad. Sci. USA*, 99: 11525–11530
- Tommasini, L., T. Schnurbusch, D. Fossati, F. Mascher and B. Keller, 2007. Association mapping of Stagonospora nodorum blotch resistance in modern European winter wheat varieties. *Theor. Appl. Genet.*, 115: 697–708
- Virk, P.S., B.V. Ford-Lloyd, M.T. Jackson, H.S. Pooni, T.P. Clemeno and H.J. Newbury, 1996. Predicting quantitative variation within rice germplasm using molecular markers. *Heredity*, 76: 296–304
- Wang, W., K. Thornton, A. Berry and M. Long, 2002. Nucleotide variation along the Drosophila melanogaster fourth chromosome. *Science*, 295: 134–137
- Weiss, K.M. and A.G. Clark, 2002. Linkage disequilibrium and mapping of human traits”. *Trends Genet.*, 18: 19–24
- Wilcox, P.L., E.C. Echt and R.D. Burdon, 2007. Gene-assisted selection: applications of association genetics for forest tree breeding”. In: Oraguzie, N.C., E.H.A. Rikkerink, S.E. Gardiner and H.N. De Silva (eds.), *Association Mapping in Plants*, pp: 211–247. Springer, New York, USA
- Wilson, L.M., S.R. Whitt, A.M. Ibanez, T.R. Rocheford, M.M. Goodman and E.S. Buckler, 2004. Dissection of maize kernel composition and starch production by candidate gene association. *Plant Cell.*, 16: 2719–2733
- Xing, Y., U. Frei, B. Schejbel, T. Asp and T. Lubberstedt, 2007. Nucleotide diversity and linkage disequilibrium in 11 expressed resistance candidate genes in *Lolium perenne*. *BMC Plant Biol.*, 7: 1–12
- Yan, W., Y. Li, H. Agrama, D. Luo, F. Gao, X. Lu and G. Ren, 2009. Association mapping of stigma and spikelet characteristics in rice (*Oryza sativa* L.). *Mol. Breed.*, 24: 277–292
- Yao, J., L. Wang, L. Liu, C. Zhao and Y. Zheng, 2009. Association mapping of agronomic traits on chromosome 2A of wheat. *Genetica*, 137: 67–75
- Yin, T.M., S.P. DiFazio, L.E. Gunter, S.S. Jawdy, W. Boerjan and G.A. Tuskan, 2004. Genetic and physical mapping of clampsora rust resistance genes in *Populus* and characterization of linkage disequilibrium and flanking genomic sequence. *New Phytol.*, 164: 95–105
- Yu, J., G. Pressoir, W.H. Briggs, I.V. Bi, M. Yamasaki, J.F. Doebley, M.D. McMullen, B.S. Gaut, D.M. Nielsen, J.B. Holland, S. Kresovich and E.S. Buckler, 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.*, 38: 203–208
- Zhang, S., C. Chen, L. Li, L. Meng, J. Singh, N. Jiang, X.W. Deng, Z.H. He and P.G. Lemaux, 2005. Evolutionary expansion, gene structure, and expression of the rice wall-associated kinase gene family. *Plant Physiol.*, 139: 1107–1124
- Zhang, J., C. Hao, Q. Ren, X. Chang, G. Liu and R. Jing, 2011. Association mapping of dynamic developmental plant height in common wheat. *Planta*, 234: 891–902
- Zhao, K., M.J. Aranzana, S. Kim, C. Lister, C. Shindo, C. Tang, C. Toomajian, H. Zheng, C. Dean, P. Marjoram and M. Nordborg, 2007. An Arabidopsis example of association mapping in structured samples. *PLoS Genet.*, 3: e4
- Zhu, Y.L., Q.J. Song, D.L. Hyten, C.P. Van Tassell, L.K. Matukumalli, D.R. Grimm, S.M. Hyatt, E.W. Fickus, N.D. Young and P.B. Cregan, 2003. Single-nucleotide polymorphisms in soybean. *Genetics*, 163: 1123–1134

(Received 03 August 2011; Accepted 05 October 2011)