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

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS BOOK CITATION INDEX INDEXED

WEB OF SCIENCE™

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

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

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

3

Association Mapping in Plant Genomes

Braulio J. Soto-Cerda[1,2] and Sylvie Cloutier[1]
[1]*Cereal Research Center, Agriculture and Agri-Food Canada*
[2]*Agriaquaculture Nutritional Genomic Center, CGNA (R10C1001)*
Genomics and Bioinformatics Unit
[1]*Canada*
[2]*Chile*

1. Introduction

One of the many goals of plant geneticists and breeders pertains to the explanation of phenotypic variation as it relates to changes in DNA sequence (Myles et al., 2009). The development of molecular markers for the detection and exploitation of DNA polymorphisms in plant systems is one of the most significant developments in the field of molecular biology and biotechnology. Linkage mapping has been a key tool for identifying the genetic basis of quantitative traits in plants. However, for linkage studies, suitable crosses, sometimes limited by low polymorphism or small population size, are required. In addition, only two alleles per locus and few recombination events are considered to estimate the genetic distances between marker loci and to identify the causative genomic regions for quantitative trait loci (QTL), thereby limiting the mapping resolution. To circumvent these limitations, linkage disequilibrium (LD) mapping or association mapping (AM) has been used extensively to dissect human diseases (Slatkin, 2008). This approach has received increased attention during the last few years. AM has the potential to identify a single polymorphism within a gene that is responsible for phenotypic differences. AM involves searching for genotype-phenotype correlations among unrelated individuals. Its high resolution is accounted for by the historical recombination accumulated in natural populations and collections of landraces, breeding materials and varieties. By exploiting broader genetic diversity, AM offers three main advantages over linkage mapping: mapping resolution, allele number and time saving in establishing a marker-trait association and its application in a breeding program (Flint-Garcia et al., 2003). Although AM presents clear advantages over linkage mapping, they are often applied in conjunction, especially to validate the associations identified by AM, thus reducing spurious associations.

The inherent nature of AM brings its own limits such as the fact that biological and evolutionary factors affect LD distribution and mapping resolution. The strength of AM relies on the analysis of common variants, which explain at most 5%-10% of the heritable component of human diseases (Asimit & Zeggini, 2010). The role of rare variants in explaining the remaining heritable variation is becoming more important. New statistical models for AM are being developed to better consider rare variants because early methods allocated most of the statistical power to higher frequency alleles.

Since most of the traits important for environmental fitness and agricultural value are quantitative in nature (Yu & Buckler, 2006), there is tremendous interest in using AM to examine them. In this chapter, we introduce the concept of linkage disequilibrium, which plays a central role in association analysis. For this reason, it is critical to understand LD measurement, its variation across the genome and how it is affected by population structure and relatedness. Recent AM studies in plants, advantages and disadvantages of AM, and its integration with other mapping methods are also reviewed and discussed. An overview of the software currently available for AM and their main characteristics is presented. Future perspectives of AM in plants, application in other emerging research areas, potential usefulness for new cultivar development and for the conservation of adaptive genetic variation are outlined.

2. Linkage disequilibrium and association mapping concepts

The terms LD and AM have often been used interchangeably in the literature. However, they present subtle differences. According to Gupta et al. (2005), AM refers to the significant association of a marker locus with a phenotype trait while LD refers to the non random association between two markers or two genes/QTLs (Figure 1). Thus, AM is actually an application of LD. In other words, two markers in LD represent a non random association between alleles, but do not necessarily correlate/associate with a particular phenotype, whereas association implies a statistical significance and refers to the covariance of a marker and a phenotype of interest. Although it lies outside the scope of this section, we would like to also clarify the difference between linkage and LD because they too are commonly 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). Although tight linkage between alleles on the same chromosome generally translate into high LD, significant LD may also exist between distant loci, and even between loci located on different chromosomes. The latter, reviewed in depth below, is the result of other forces such as selection, mutation, mating system, population structure, etc.

Both QTL and AM approaches are therefore based on LD between molecular markers and functional loci. In QTL mapping, LD is generated by the mating design while in AM, LD is a reflection of the germplasm collection under study (Stich & Melchinger, 2010). In a mapping population, LD is influenced only by recombination in the absence of segregation distortion. In AM, LD may also be influenced by other forces such as those mentioned above as well as by recombination.

The concept of LD was first described by Jennings in 1917, and its quantification (D) was developed by Lewontin in 1964 (Abdurakhmonov & Abdukarimov, 2008). The simplified explanation of the commonly used LD measure, D or D' (standardized version of D), is the difference between the observed gametic frequencies of haplotypes and the expected gametic frequencies of haplotype under linkage equilibrium.

$$D = P_{AB} - P_A P_B \quad (1)$$

Where P_{AB} is the frequency of gametes carrying allele A and B at two loci; P_A and P_B are the product of the frequencies of the allele A and B, respectively. In the absence of other forces, recombination through random mating breaks down the LD with $D_t = D_0 (1 - r)^t$, where D_t is the remaining LD between two loci after t generations of random mating from the original D_0 (Zhu et al., 2008).

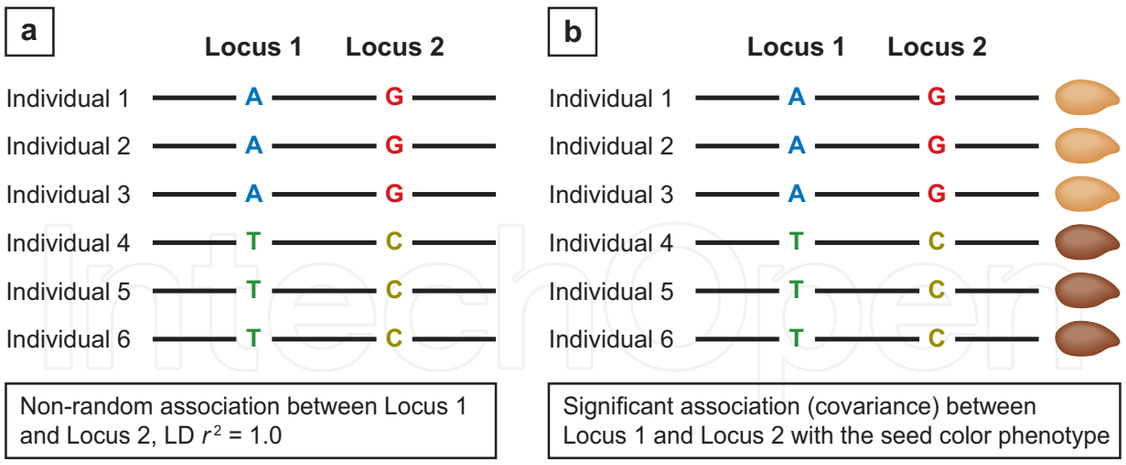


Fig. 1. Principles of linkage disequilibrium and association mapping. a. Linkage disequilibrium. Locus 1 and Locus 2 present an unusual pattern of association between alleles A-G and T-C, which deviate from Hardy-Weinberg expectations, but without any statistical correlation with a phenotype. b. Association mapping. Locus 1 and Locus 2 are in LD. Significant covariance with the seed colour phenotype is considered evidence of association.

Several statistics have been proposed for LD, and these measurements generally differ in how they are affected by marginal allele frequencies and sample sizes. Here, we introduce the two most utilized statistics for LD. Both D' (Lewontin, 1964) and r^2, the square of the correlation coefficient between two loci (Hill & Robertson, 1968), reflect different aspects of LD and perform differently under various conditions. D' only reflects the recombinational history and is therefore a more accurate statistic for estimating recombination differences, whereas r^2 summarizes both recombinational and mutational history (Flint-Garcia et al., 2003). For two biallelic loci, D' and r^2 have the following formula:

$$D' = |D| \,/\, D_{max} \tag{2}$$

$$D_{max} = \min (P_A P_b, P_a P_B) \text{ if } D > 0;$$

$$D_{max} = \min (P_A P_B, P_a P_b) \text{ if } D < 0$$

$$r^2 = D^2 \,/\, P_A P_a P_B P_b \tag{3}$$

D is limited because its range is determined by allele frequencies. D' was developed to partially normalize D with respect to the maximum value possible for the allele frequencies and give it a range between 0 and 1 (Zhu et al., 2008). The r^2 statistic has an expectation of $1/(1+4Nc)$, where N is the effective population size and c is the recombination rate, and it also varies between 0 and 1 (Hill & Robertson, 1968).

Choosing the appropriate LD statistics depends on the objective of the study. Most studies on LD in animal populations used D' to measure population-wide LD of microsatellite data (Du et al., 2007). However, D' is inflated by small sample size and low allele frequencies; therefore, intermediate values of D' are unsafe for comparative analyses of different studies and should be verified with r^2 before being used for quantification of the

extent of LD (Oraguzie et al., 2007). Although r^2 is still considered to be allele frequency dependent, the bias due to allele frequency is considerably smaller than in D' (Ardlie et al., 2002). Currently, most LD mapping studies in plants use r^2 for LD quantification because it also provides information about the correlation between markers and QTL of interest (Flint-Garcia et al., 2003; Gupta et al., 2005). Typically, r^2 values of 0.1 or 0.2 are often considered the minimum thresholds for significant association between pairs of loci and to describe the maximum genetic or physical distance at which LD is significant (Zhu et al., 2008).

3. Visualization and statistical significance of LD

Since D' and r^2 are pairwise measurements between polymorphic loci, it is difficult to obtain summary statistics of LD across a region (Gupta et al., 2005). There are two common ways to visualize the extent of LD and the genomic regions or haplotype blocks found to be in significant LD. LD scatter plots are used to estimate the rate at which LD declines with genetic or physical distance (Figure 2a). An average genome-wide decay of LD can be estimated by plotting LD values, from a data set covering an entire genome, against distance. Alternatively, the extent of LD can be estimated for a particular region carrying a gene/QTL of interest previously identified by linkage mapping. These scatter plots are useful to determine the average effective distance threshold above which significant LD (commonly 0.5 for D' and 0.1 for r^2) is expected based on the curve of a nonlinear logarithmic trend drawn through the data points of the scatter plot (Breseghello & Sorrells, 2006). Disequilibrium matrices or LD heat maps are also very useful for visualizing the linear arrangement of LD between polymorphic sites within a short physical distance such as a gene, along an entire chromosome or across the whole genome (Figure 2b) (Flint-Garcia et al., 2003). LD heat maps are colour-coded triangular plots where the diagonal represents ordered loci and the different intensity coloured pixels depict significant pairwise LD level expressed as D' or r^2. Blocks of high intensity pixels afford an easy visualization of loci in significant LD. In this figure, the larger the blue blocks of haplotypes along the diagonal of the triangular plot, the higher the level and extent of LD between adjacent loci in the blocks, meaning that there has been either limited or no recombination since the LD block formation (Abdurakhmonov & Abdukarimov, 2008). These graphical representations enable us to determine the optimum number of markers to detect significant marker-trait associations and the resolution at which a QTL can be mapped. Because LD estimation based on D' or r^2 can be sensitive to marker density, highly saturated and representative linkage groups are ideal for LD calculations.

The statistical significance of LD is typically determined using a χ^2 test of a 2 × 2 contingency table. A p-value threshold of 0.05 is often used to declare lack of independence of alleles at two loci, thus suggesting association (Gupta et al., 2005). From a 2 × 2 contingency table, the probability (P) of independence of alleles at two loci is generally calculated through a Fisher's exact test (Fisher, 1935; as cited in Gupta et al., 2005). Statistically significant LD can also be calculated using a multifactorial permutation analysis to compare sites with more than two alleles such as microsatellite markers. These statistical methods are implemented in software such as PowerMarker (Liu & Muse, 2005) and TASSEL (Trait Analysis by aSSociation Evolution and Linkage) (Bradbury et al., 2007).

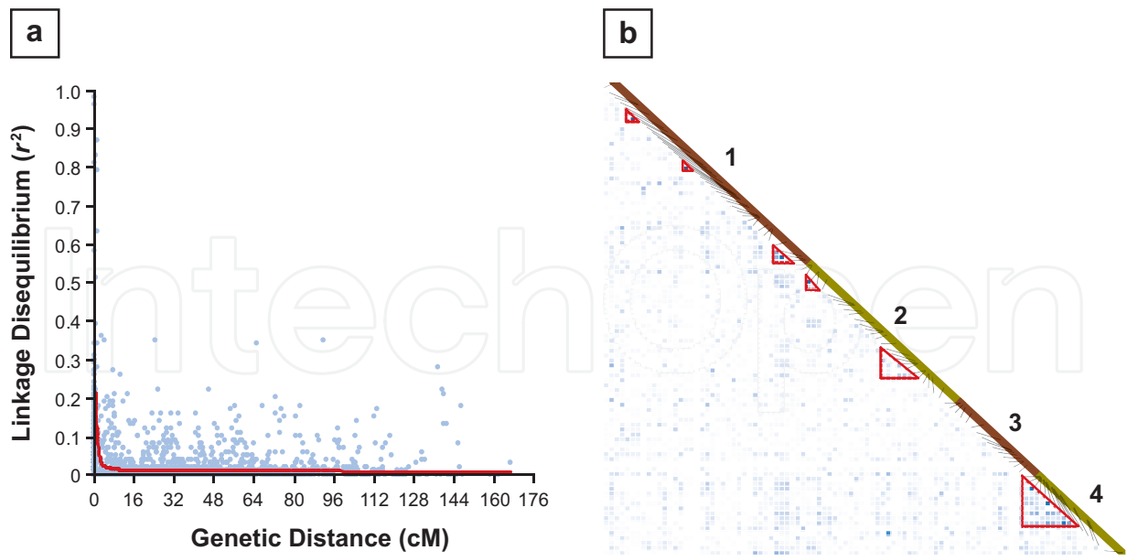


Fig. 2. Visualization of linkage disequilibrium in flax (*Linum usitatissimum* L.). a. Scatter plot of LD decay (r^2) against genetic distance (cM), representing a measure of an average genome-wide LD. b. Heatmap of LD variation between pairwise polymorphic loci of four linkage groups. Blocks in significant LD are highlighted by red triangles. LD distribution is heterogeneous within and between linkage groups.

4. LD variation as an effect of biological factors

4.1 Recombination

Several biological factors influence LD strength and its distribution across genomes. Many regions of the human genome display rates of recombination that differ significantly from the genome average recombination rate of 1 cM/Mb (Arnheim et al., 2003). These regions have been called "hotspots" and "coldspots" for high and low recombination rates, respectively. LD is strongly influenced by localized recombination rate and is correlated with other associated factors such as GC content and gene density (Dawson et al., 2002). In principle, local sequence features can affect LD directly and indirectly. For example, GC-rich sequences may be associated with higher rates of recombination and/or mutation, two phenomena that could directly lower surrounding levels of LD. Furthermore, in some protein-coding sequences, changes created by recombination or mutation may affect the fitness of an individual, and these sequences could be indirectly associated with unique patterns of LD as a consequence of natural selection (Smith et al., 2005).

Because LD is broken down by recombination, and recombination is not distributed homogeneously across the genome, blocks of LD are expected. Also, differences in LD between micro chromosomes and macro chromosomes have been reported (Stapley et al., 2010) as well as intra-chromosomal variation, where centromeric regions showed higher levels of LD. Teo et al. (2009) conducted a comprehensive analysis of genomic regions with different patterns of LD to unravel the consequences of this patterning for AM in human populations. Plant genomes have revealed similar general conclusions with regards to LD distribution. Inter-chromosomal LD variation has been reported in barley (*Hordeum vulgare*), maize (*Zea mays*), tomato (*Solanum lycopersicum*) and bread wheat (*Triticum aestivum*)

(Malysheva-Otto et al., 2006; Robbins et al., 2011; Yan et al., 2009; Zhang et al. 2010), where it varied between less than 1 cM to more than 30 cM ($r^2 > 0.1$). As a consequence, investigation of LD variation at the genome and chromosome scale to accurately estimate marker density for each chromosome is required to provide insights to the most cost-effective AM approach.

4.2 Mating system

The mating system has profound effects on LD (Myles et al., 2009). Selfing reduces opportunities for effective recombination because individuals are more likely to be homozygous than in outcrossing species (Flint-Garcia et al., 2003). In self-pollinated species such as rice (*Oryza sativa*), Arabidopsis (*Arabidopsis thaliana*) and wheat (*Triticum aestivum*) (Garris et al., 2005; Nordborg, 2000; Zhang et al. 2010), LD extends much further as compared to outcrossing species such as maize (*Zea mays*), grapevine (*Vitis vinifera*) and rye (*Secale cereale*) (Li et al., 2011b; Myles et al., 2009; Tenaillon et al., 2001). As a result, genetic polymorphisms tend to remain correlated, and LD is expected to be maintained over long genetic or physical distances (Gaut & Long, 2003). However, because LD declines more rapidly in outcrossing plant species than self-pollinated plants, a higher resolution is expected, enabling more accurate fine mapping and potentially facilitating the cloning of candidate genes. A detailed review of LD decay in self-pollinated and outcrossing species can be found in Flint-Garcia et al. (2003) and Abdurakhmonov & Abdukarimov (2008).

4.3 Germplasm

The germplasm plays a key role in LD variation because the extent of LD is influenced by the level of genetic diversity captured by the population under consideration. In general, the larger the genetic variation, the faster the LD decay, a direct consequence of the broader historical recombination. The population sample effect is evident in maize (*Zea mays*) where LD decays within 1 kb in landraces, approximately doubles (~ 2kb) in diverse inbred lines and can extend up to several hundred kb in commercial elite inbred lines (Jung et al., 2004). Tenaillon et al. (2001) investigated sequence diversity at 21 loci on chromosome 1 in a diverse group of maize germplasm, including exotic landraces and US accessions. An average LD decay was determined to occur within 400 bp ($r^2 = 0.2$), but extended up to 1000 bp in a group of US inbred lines. In Michigan local Arabidopsis populations, LD decay varied within 50 kb up to 50-100 cM. The latter was explained as a genetic bottleneck or founder effect, which reduced dramatically the genetic variation (Nordborg et al., 2002). In cotton (*Gossypium hirsutum*), the genome-wide average LD ($r^2 \leq 0.1$) declined to 10 cM in landraces, but was up to 30 cM in varieties (Abdurakhmonov et al., 2008). Myles et al. (2011) studied LD variation in over 1000 samples of domesticated grape (*Vitis vinifera*) and its wild relatives, reporting a rapid LD decay, even greater than in maize, as result of a weak domestication bottleneck followed by thousands of years of widespread vegetative propagation.

Estimates of genome-wide average LD decay may not reflect LD patterns between different populations of the same species. Each of these populations should be explored independently for the extent of LD in order to conduct successful association mapping studies (Abdurakhmonov & Abdukarimov, 2008). Taking into account these three important biological factors, an obvious question is whether an increased or decreased level of LD is

favourable in AM? Populations with either rapid or slow LD decay can be useful in AM, depending on the purposes of the study. Thus, populations with narrow genetic diversity and long extent of LD are amenable to coarse mapping with fewer markers requiring fine mapping in more genetically diverse populations, assuming that the causal genetic factors are sufficiently similar across different germplasm groups.

5. LD variation as effect of evolutionary factors

5.1 Selection

Initial interest in LD arose from questions surrounding the *modus operandi* of natural selection. Simply stated, if alleles at two loci are in LD and they both affect positively reproductive fitness, the response to selection at one locus might be accelerated by selection affecting the other (Slatkin, 2008). Thus, positive selection will increase LD between and in the vicinity of the selected loci, a phenomenon known as genetic hitchhiking (Maynard Smith & Haigh, 1974; as cited in Slatkin, 2008). Even if the second locus is selectively neutral, the selection applied over the first will increase LD between them. The LD level between the two loci will remain constant over time depending on the genetic distance, the recombination rate and the effective population size (N). In contrast, if both loci are maintained by balancing selection, then LD can persist indefinitely (Lewontin, 1964). Nonetheless, LD should be higher in loci affected by positive selection because a strong positive selection limits genetic diversity as opposed to a balancing selection which tends to maintain or increase polymorphism. In general, disease resistance genes in plants (*R*-genes) are affected by balancing selection with low intragenic LD and rapid decay (Yin et al., 2004), which could facilitate fine mapping of disease resistance genes providing high marker saturation. Artificial selection also has dramatic effects on LD. Mosaics of large LD blocks are observed, especially in regions carrying agronomic-related genes. Domestication bottlenecks followed by strong selection for specific environments and end-use traits have modified the genome architecture in many crops reducing genetic diversity and creating population structure, which may be the main factor affecting the power of AM.

5.2 Population structure

Selection affects the genome and LD in locus-specific manner. In contrast, population structure affects LD throughout the genome. Consequently, genome-wide patterns of LD can help to understand the history of changes in populations (Slatkin, 2008). However, the power of AM can be strongly reduced as a consequence of population structure (Balding, 2006). Population structure occurs from the unequal distribution of alleles among subpopulations of different ancestries. When these subgroups are sampled to construct a panel of lines for AM, the intentional or unintentional mixing of individuals with different allele frequencies creates LD. Significant LD between unlinked loci results in false-positive associations between a marker and a trait. The effect is obvious in the following case. Suppose that one subpopulation is fixed for *A* and *B* alleles at two loci whereas another is fixed for *a* and *b*. Any mixture of individuals from the two subpopulations would contain only *AB* and *ab* haplotypes, implying that they are in perfect LD, when in fact there is no LD in either subpopulation (Slatkin, 2008). By definition, polymorphisms at two or more loci must exist to estimate the level of LD. In the above example, both loci are monomorphic in their respective subpopulations. However, when individuals are mixed, in the newly created artificial single population, false

polymorphisms and consequently significant but spurious LD is observed. Thornsberry et al., (2001) reported significant associations between polymorphisms at the maize *Dwarf8* gene and variation in flowering time, but they also stated that up to 80% of the false positive associations resulted from population structure. The occurrence of spurious associations is markedly higher in adaptation-related genes because they show positive correlations with the environmental variables under which they have evolved, and, as a result, the genomic regions carrying these genes could present stronger population differentiation. Several statistical models take into account the potential effect of population structure. Commonly used algorithms are those of Pritchard & Rosenberg (1999) implemented in the software STRUCTURE (Hubisz et al., 2009; Pritchard et al., 2000). Other methods are based on Principal Component Analysis (PCA) (Price et al., 2006), and Principal Coordinate Analysis and Modal Clustering (PCoA-MC) (Reeves & Richards, 2009).

5.3 Genetic drift, population bottleneck and gene flow

The effect of genetic drift in a small population results in the consistent loss of rare allelic combinations which increases LD level (Flint-Garcia et al., 2003). Genetic drift can create LD between closely linked loci. The effect is similar to taking a small sample from a large population. Even if two loci are in linkage equilibrium, sampling only few individuals can create LD (Slatkin, 2008).

LD can also be created in populations that have experienced a reduction in size (called a bottleneck) with accompanying extreme genetic drift (Dunning et al., 2000; as cited in Flint-Garcia et al., 2003). After a bottleneck, some haplotypes will be lost; generally resulting in increased LD. Subsequent bottlenecks will further contribute to augment LD by increasing the effect of genetic drift. Colonizing species undergo repeated bottlenecks, and many models of the history of hominids assume the occurrence of a bottleneck when modern humans first left Africa (Noonan et al., 2006). Several studies of humans have argued that long distance LD in humans is the result of this early bottleneck in human history (Schmegner et al., 2005). In plants, comparisons with wild ancestors indicate that, in maize, approximately 80% of the allele richness has been lost as a consequence of domestication bottlenecks (Wright & Gaut, 2005) while this number is 40-50% in sunflower (Liu & Burke, 2006) and 10-20% in rice (Zhu et al., 2007). Gene flow introduces new individuals or gametes with different ancestries and allele frequencies among populations. If selection maintains differences in allele frequencies at two or more loci among subpopulations, LD in each subpopulation will persist (Slatkin, 1975; as cited in Slatkin, 2008), but generally when random mating and recombination take place, LD caused by gene flow eventually breaks down.

Factors such as genetic drift, population bottlenecks and gene flow can contribute to generating artificial LD and negatively impact the ability to use LD in AM for the precise localization of QTL. In general, any biological or evolutionary forces that contribute to an increase of LD beyond that expected by chance in an "ideal" population will result in false-positive associations (Gaut & Long, 2003).

6. Approaches for AM

Many methodologies have been developed and are widely used for AM in humans (Schulze & McMahon, 2002), and several are perfectly applicable without change or with case-to-case

modifications for a wide range of organisms, including plants. The methods to study marker-trait association using LD may differ for discrete and quantitative traits (Nielsen & Zaykin, 2001). Here, we will examine several approaches: Multiparent Advanced Generation Intercross (MAGIC), Case-control (CC), Transmission Disequilibrium Test (TDT) and other approaches that incorporate corrections for population structure such as genomic control (GC) and structured association (SA).

6.1 Multiparent Advanced Generation Intercross (MAGIC)

MAGIC is an extension of the advanced intercross method in which an intermated mapping population is created from multiple founder lines. A Recombinant Inbred Line (RIL) population is created from multiple founder lines, in which the genome of the founders are first mixed by several rounds of mating, and subsequently inbred to generate a stable panel of inbred lines. The larger number of parental accessions increases the allelic and phenotypic diversity over traditional RILs, potentially increasing the number of QTL that segregate in the population. The successive rounds of recombination cause LD to decay, thereby increasing the precision of QTL location (Mackay & Powell, 2007). In both crops and animals, the MAGIC design has the ability to capture the majority of the variation available in the gene pool. Although it might take several years before these populations are suitable for fine mapping, they are relatively inexpensive to develop and their value as mapping resources increases with each generation (Mackay & Powell, 2007). In plants, MAGIC can be used to combine coarse mapping with low marker densities on lines derived from an early generation, with fine mapping using lines derived from a more advanced generation and a higher marker density. Regardless of the generation used, LD decay remains the critical factor determining the mapping resolution.

6.2 Case-control (CC)

The classical methodology and design of AM is the "case and control" (CC) approach. If a mutation increases disease susceptibility, then we can expect it to be more frequent among affected individuals (cases) than among unaffected individuals (controls). The essential idea behind CC-based AM is that markers close to the disease mutation may also have allele frequency differences between cases and controls if there is LD between the marker locus and the "susceptibility" mutations (Schulze & McMahon, 2002). For accurate mapping, this design requires an equal number of unrelated and unstructured *case-control* samples. The Pearson χ^2 test, Fisher's exact test or Yates continuity correction can be used to compare allele frequencies and detect association between a phenotype and a marker (Abdurakhmonov & Abdukarimov, 2008). The CC tests are sensitive to overall population LD between a marker and a locus affecting the trait. As previously discussed, LD can exist between unlinked loci, meaning that strong marker-trait association is not necessarily evidence for physical proximity between a marker and the gene affecting the phenotype. As a consequence, the CC approach is highly sensitive to population structure (Schulze & McMahon, 2002). To efficiently eliminate the confounding effects caused by population structure, Spielman et al. (1993) developed the Transmission Disequilibrium Test (TDT).

6.3 Transmission Disequilibrium Test (TDT)

The ability to map QTL in collections of breeding lines, landraces or samples from natural populations has merit. In these populations, LD often decays more rapidly than in

controlled crosses, enabling fine mapping. The challenge is to distinguish the effects of population subdivision from LD caused by linkage (syntenic LD). A robust method to test for this partitioning is the TDT (Spielman et al., 1993) that permits the detection of linkage in the presence of disequilibrium. Neither linkage alone nor disequilibrium alone (non syntenic LD) will generate a positive result in a TDT. As a consequence, the TDT is a robust method to control false positives (Mackay & Powell, 2007). In brief, TDT compares the transmission versus the non transmission of alleles to the offspring using a χ^2 test, assuming a linkage between a marker and a trait. The TDT design requires genotyping of markers from three individuals: one heterozygous parent, one homozygous parent and one affected offspring. In the absence of linkage between QTL and marker, the expected ratio of transmission to non transmission is 1:1 (Nielsen & Zaykin, 2001). In the presence of linkage, it is distorted to an extent that depends on the strength of LD between the marker and the QTL. In addition, the power of the association will depend on the effectiveness of selection of extreme progeny in driving segregation away from expectation (Mackay & Powell, 2007).

The initial TDT approach did not address the cases of multiallelic markers, multiple markers, missing parental information, large pedigrees and complex quantitative traits (Schulze & McMahon, 2002). A variety of extensions of the TDT approach have been developed and applied to resolve multiallelic marker issues (i.e., GTDT, ETDT, MCTm); reviewed by Schulze & McMahon (2002).

In crops, parental and progeny lines are often separated by several generations of gametogenesis rather than one, as is often the case of human studies. For this reason, the TDT, while still valid, may be less robust because the breeding process may result in increased segregation distortion (Mackay & Powell, 2007).

6.4 Other approaches

Population structure arising from recent migration, population admixture and artificial selection will generate non syntenic LD. Assuming that such population structure has a similar effect on all loci, a random set of markers can be used to statistically assess the extent with which population structure is responsible for non syntenic LD (Stich & Melchinger, 2010). This is the basis of genomic control (GC). For example, for a case-control analysis of candidate genes, the GC approach computes χ^2 test statistics for independence for both null (random) and candidate loci. An average χ^2 of null loci greater than 1.0 indicates the presence of significant structure. By using the magnitude of the χ^2 test observed at the null loci, a multiplier is derived to adjust the critical value for significance tests for candidate loci (Mackay & Powell, 2007). By contrast, structure association (SA) analysis developed by Pritchard et al. (2000), first uses a set of random markers to estimate population structure (*Q*-matrix), and then incorporates this estimate into a general linear model (GLM) analysis which enables correction for false associations. Yu et al. (2006) developed a new methodology, the mixed linear model (MLM), which incorporates both population structure and familial relatedness or so-called "kinship" (*K*-matrix). To perform MLM: (1) a *Q*-matrix is generated using for example, STRUCTURE; (2) the pairwise relatedness coefficients between individuals of a germplasm collection (*K*-matrix) is estimated using for example, SpaGeDi software (Hardy & Vekemans, 2002); and (3) both *Q*- and *K*-matrices are used in AM to control spurious associations. Studies conducted in human, Arabidopsis and bread

wheat (Raman et al., 2010; Yu et al., 2006; Zhao et al., 2007a) have demonstrated the effectiveness of the MLM approach over the GLM.

Another type of mixed model used in AM incorporates PCA instead of the *Q*-matrix. Promising results as a fast and effective way to identify population structure have been reported (Price et al., 2006). The PCA-based MLM model is computationally effective as compared to the *Q*-matrix estimated from STRUCTURE. Also, STRUCTURE has been found to overestimate the "true" number of subpopulations under particular scenarios (Evanno et al., 2005).

7. AM studies in plants

Some of the first LD mapping studies in plants were done in maize (*Zea mays*) (Bar-Hen et al., 1995), rice (*Oryza sativa*) (Virk et al., 1996) and oat (*Avena sativa*) (Beer et al., 1997). Bar-Hen et al. (1995) and Virk et al. (1996) predicted the association of quantitative traits using RAPD and isozymes markers, respectively. Beer et al. (1997) associated 13 QTL with RFLP loci using 64 oat varieties and landraces. In these studies, a low number of genome-wide distributed markers were assessed without considering the population structure. The first empirical candidate gene association taking into account background molecular markers to correct for population structure was performed in maize looking at the D8 locus and its association with flowering time (Pritchard, 2001). In Arabidopsis, most of the AM studies focused on providing proof of concept, identification of QTL involved in adaptation and detection of additional alleles to supplement other mutagenesis approaches (Ersoz et al., 2007). Aranzana et al. (2005) performed the first attempt at a genome wide association study (GWAS) in Arabidopsis, reporting previously known flowering time and three known pathogen-resistance genes. GWAS refers to the use of many markers that span an entire genome to identify functional common variants in LD with at least one of the genotyped markers. Numerous research papers focusing on LD and AM have since been published on more than a dozen plant species. These studies have been reviewed by Gupta et al. (2005) and more recently by Zhu et al. (2008).

In the last five years, plant AM studies have expanded because of advances in sequencing technologies which enable more efficient and cost-effective development of a large number of molecular markers such as Single Nucleotide Polymorphisms (SNPs). In Arabidopsis, new studies have been carried out aiming to dissect downy mildew resistance genes and climate-sensitive QTL, with special efforts focused on the understanding of adaptative variation (Li et al., 2010; Nemri et al., 2010). The first applied a CG approach, and the second a GWAS based on no fewer than 213,497 SNPs. In maize, recent studies dissected the quantitative genetic nature of the northern leaf blight (NLB) resistance, southern leaf blight (SLB) resistance and leaf architecture, scanning the genome using ~ 1.6 million SNPs (Kump et al., 2011; Poland et al., 2011; Tian et al., 2011). Poland et al. (2011) identified several loci with small additive effects carrying candidate genes related to plant defense, including receptor-like kinase genes. Kump et al. (2011), from the same research group, identified 32 QTL with predominantly small additive effects related to SLB resistance. Similarly, Tian et al. (2011) demonstrated that the genetic architecture of leaf traits is dominated by small effects and that the *liguleless* genes have contributed to more upright leaves. Currently, whole genome scanning has moved beyond Arabidopsis and maize to other species such as rice and barley. Huang et al. (2010) uncovered the genetic basis of 14 rice agronomic traits

based on ~ 3.6 million SNPs. The loci identified through GWAS explained ~ 36% of the phenotypic variance, on average. In barley, GWAS of 15 morphological traits identified one putative anthocyanin pathway gene, *HvbHLH1*, carrying a deletion resulting in a premature stop codon and which was diagnostic for the absence of anthocyanin in the germplasm studied (Cockram et al., 2010). Efforts towards understanding adaptation-related genes have been undertaken in wheat. Raman et al. (2010) applied GWAS in order to identify genetic factors associated with aluminium resistance, one of the most restrictive abiotic stresses on acid soils worldwide. The study confirmed previously identified loci and identified putative novel ones. Subsequently, Rousset et al. (2011) studied the genetic nature of flowering time in wheat to investigate the effect of candidate genes on flowering time. The *Vrn-3* gene explained a high percentage of the phenotypic variation of earliness followed to a lesser extent by *Vrn-1*, *Hd-1* and *Gigantea* (*GI*). In *Brassica napus*, several seed oil related loci were identified, with a few corresponding to previously reported genomic regions associated with oil variation (Zou et al., 2010). In tetraploid alfalfa (*Medicago sativa*), 15 SSR markers showed strong association with yield in different environments (Li et al., 2011a). In sugar beet (*Beta vulgaris*), genetic variation of six agronomic traits was dissected using GWAS, identifying several QTL with major effects and others with epistatic effects (Würschum et al., 2011). Thus, LD mapping, considered a few years ago as an emerging tool in plant genomics, has recently been shown to be a powerful method to dissect complex traits in crops. Table 1 summarizes these and other recently published AM studies in plants. Earlier publications are summarized elsewhere (Gupta et al., 2005; Zhu et al., 2008).

8. Benefits and limitations of AM

The potential high resolution in localizing a QTL controlling a trait of interest is the primary advantage of AM as compared to linkage mapping (Figure 3). AM has the potential to identify more and superior alleles and to provide detailed marker data in a large number of lines which could be of immediate application in breeding (Yu & Buckler, 2006). Furthermore, AM uses breeding populations including diverse and important materials in which the most relevant genes should be segregating. Complex interactions (epistasis) between alleles at several loci and genes of small effects can be identified, pinpointing the superior individuals in a breeding population (Tian et al., 2011). Sample size and structure do not need to be as large as for linkage studies to obtain similar power of detection. Finally, AM has the potential not only to identify and map QTL but also to identify causal polymorphisms within a gene that are responsible for the difference between two phenotypes (Palaisa et al., 2003).

AM suffers from some limitations such as when the trait under consideration is strongly associated with population structure. Most traits under local adaptation or in balancing selection in different populations may be thus affected (Stich & Melchinger, 2010). When statistical methods to correct for population structure are applied, the differences between subpopulations are disregarded when searching for marker-trait associations. Therefore, all polymorphisms responsible for the phenotypic differences between subpopulations remain undetected, thus underpowering AM. LD mapping often requires a large number of markers for genotyping in GWAS. The number of markers depends in large part on the genome size and the expected LD decay; linkage mapping generally requires fewer markers to detect significant QTL. A high density of markers can only be achieved through the

Species	Germplasm	Trait	Marker system	Reference
Arabidopsis	Diverse accessions	Flowering time/pathogen resistance	Sequences	Aranzana et al. (2005)
	Diverse accessions	Multiple traits	SSRs/SNPs	Ersoz et al. (2007)
	Natural accessions	Flowering time	SNPs	Brachi et al. (2010)
	Diverse accessions	Climate-sensitive QTL	SNPs	Li et al. (2010)
	Landraces	Downy mildew	SNPs	Nemri et al. (2010)
Maize	Inbred lines	Aluminum tolerance	SNPs	Krill et al. (2010)
	Inbred lines	Drought tolerance	SNPs	Lu et al. (2010)
	Inbred lines	Northern leaf blight	SNPs	Poland et al. (2011)
	Inbred lines	Southern leaf blight	SNPs	Kump et al. (2011)
	Inbred lines	Leaf architecture	SNPs	Tian et al. (2011)
Teosinte	Landraces	Domestication-related genes	SNPs	Weber et al. (2009)
Wheat	Cultivars	Kernel size, milling quality	SSRs	Breseghello & Sorrells, (2006)
	Diverse accessions	Aluminum resistance	DArT	Raman et al. (2010)
	Breeding lines	Stem rust resistance	DArT	Yu et al. (2011)
	Diverse accessions	Flowering time	SNPs	Rousset et al. (2011)
Barley	Inbred lines	Growth habit	SNPs	Rostoks et al. (2006)
	Cultivars	Anthocyanin pigmentation	SNPs	Cockram et al. (2010)
	Breeding lines	Winterhardiness	SNPs	Von Zitzewitz et al. (2011)
Oat	Diverse cultivars	Agronomic and kernel quality traits	AFLPs	Achleitner et al. (2008)
Rice	Landraces	Heading date, plant height and panicle length	SSRs	Wen et al. (2009)
	Landraces	Multiple agronomic traits	SNPs	Huang et al. (2010)
Canola	Diverse accessions	Leaf traits, flowering time and phytate content	AFLPs	Zhao et al. (2007b)
	Diverse accessions	Oil content	SSRs	Zou et al. (2010)
Soybean	Breeding lines	Iron deficiency chlorosis	SSRs	Wang et al. (2008)
Cotton	Diverse cultivars	Fiber quality	SSRs	Abdurakhmonov et al. (2009)
Peanut	Diverse accessions	Seed quality traits	SSRs-SNPs	Wang et al. (2011)
Sugar beet	Inbred lines	Sugar content and yield	SSRs	Stich et al. (2008)
	Inbred lines	Multiple traits	SNPs	Würschum et al. (2011)
Alfalfa	Cultivars	Biomass yield and stem composition	SSRs	Li et al. (2011a)

SNPs: Single Nucleotide Polymorphisms; SSRs: Simple Sequence Repeats; DArT: Diversity Arrays Technology; AFLPs: Amplified Fragment Length Polymorphisms.

Table 1. Association mapping studies in plants.

development of an integrated genotyping by sequencing (GBS) platform. Thus, the analysis of cost-benefit must be conducted in the light of the real impacts that such investments will have in the future market appreciation of that plant species. Alternative approaches such as linkage mapping and CG could be feasible for other studied traits. The power of AM to detect an association is influenced by allele frequency distribution at the functional polymorphism level. The results of empirical studies suggest that a high percentage of alleles are rare (Myles et al., 2009). Rare alleles cannot be evaluated adequately because, by definition, they are present in too few individuals and consequently lack resolution power. As a consequence, an important piece of heritability remains undetected. For such rare alleles, linkage mapping may be used because correlation between population structure and phenotypes can be broken, and allele frequencies can be inflated to enhance the power of mapping (Stich & Melchinger, 2010). In this regard, several studies have combined linkage mapping and LD mapping, a methodology known as "nested association mapping", which reduces spurious associations caused by population structure, particularly for traits strongly affected by local geographic patterns (Brachi et al., 2010; Poland et al., 2011). With the growing interest in finding the missing heritability not accounted for by common alleles (Asimit & Zeggini, 2010), several new association analysis methods for rare variants are being proposed, with some important advances in complex trait dissection (Li & Leal, 2008).

9. Computer programs for AM

A variety of software packages are available for AM, and it can be inferred from the previous sections that LD studies are computationally demanding. Thus, newer and more powerful programs are constantly under development. TASSEL is a commonly used software for LD mapping in plants, frequently updated with newly developed methods. Current examples include the GLM and the multiple regression models combined with the estimates for false discovery rate. TASSEL can also be used for calculation and graphical display of LD statistics, analysis of population structure using PCA, and tree plots of genetic distance. Although TASSEL can handle both SSR and SNP markers, the latest version only accepts SNPs. For SSR analysis, users must continue with TASSEL v. 2.1. Alternatively, GenStat offers traditional statistical analyses as well as linkage and AM analyses for SSRs.

GenStat performs structure analysis based on PCA, LD decay and single trait association analysis using PCA-based MLM. Version 14 was recently released and can be downloaded for non profit purposes from http://www.vsni.co.uk/2011/asides/genstat-14-released/. Gupta et al. (2005) and Excoffier & Heckel (2006) comprehensively reviewed the most common software for population genetics and LD mapping analyses but the majority of them can only handle a few thousand marker loci. Progress in sequencing technologies has solved the past issue of genotyping large populations with high marker densities and software development has also moved quickly. Nowadays, the main issue is the time required for processing large data sets and the availability of powerful statistical models to adjust for multiple testing. JMP Genomics v.5 is a Windows based program that offers several solutions for handling large SNP data sets (http://www.jmp.com/software/genomics). Among its main characteristics, JMP Genomics is capable of handling data sets as large as 1.5 million SNPs for 15,000 samples on a 32-bit desktop work station using CG or GWA. It also corrects for relatedness and population structure using association tests, and calculates identical by descent (IBD), identical by state (IBS) and allele-sharing individual relationship matrices. Interactive triangular plots

and zooming features permit visualization of LD blocks. Association between SNPs and multiple traits can be tested separately or jointly, while adjusting for covariates. JMP Genomics 5 also simplifies the analysis of rare and common variants, and includes features for high quality graphs and figures.

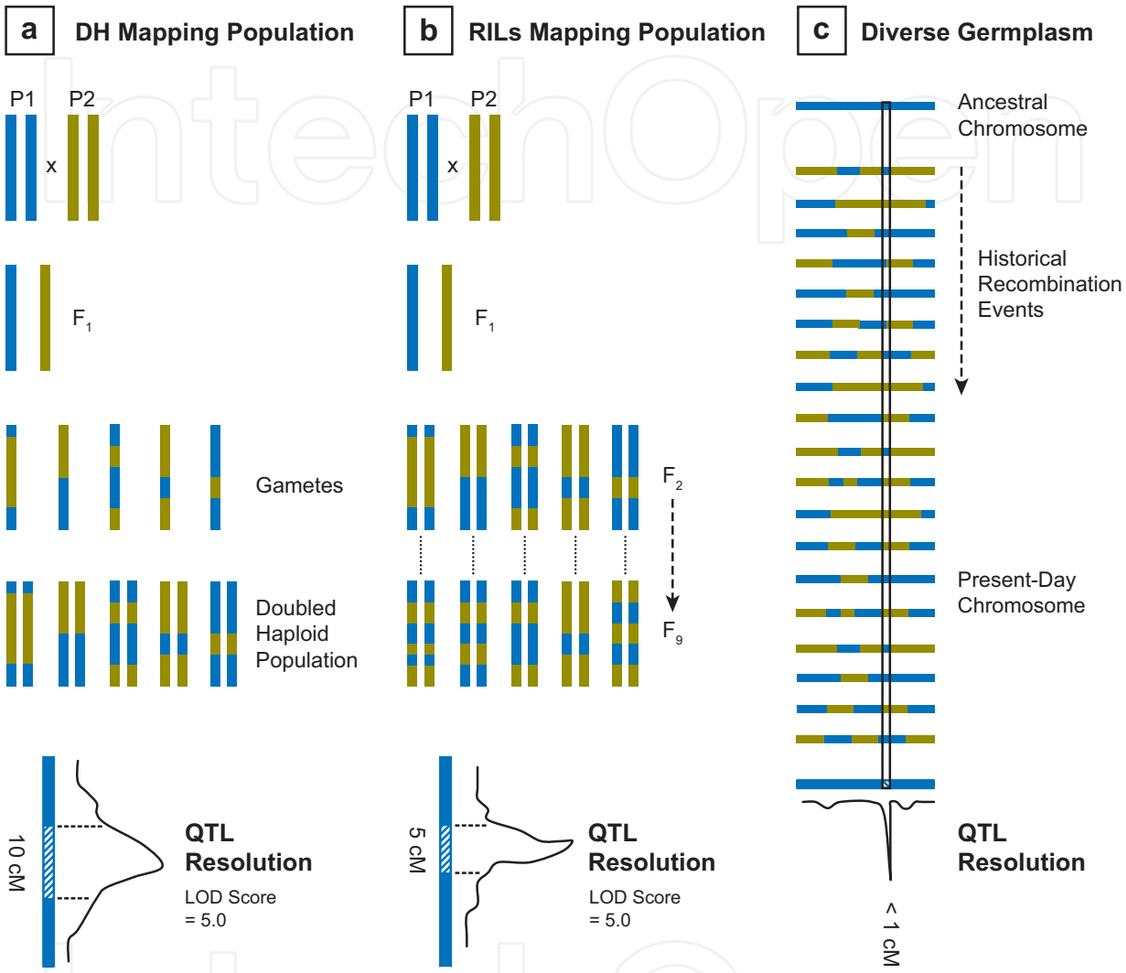


Fig. 3. Comparison of mapping resolution between linkage mapping and AM. a. A Doubled Haploid (DH) mapping population. b. A Recombinant Inbred Line (RIL) mapping population. c. A collection of diverse germplasm. **a** and **b** present low QTL resolution as a consequence of few meiosis events accumulated; **c** presents a high QTL resolution because a larger number of recombination events have accumulated during the population history.

Similar applications can be found in GenAMap software, which incorporates visualization strategies for structured AM (http://cogito-b.ml.cmu.edu/genamap/). It has a processing capacity of 1 million SNPs in approximately 1 hour. The analysis is performed on a remote cluster complete with complex parallelization schemes to optimize run-time efficiency. GenAMap gives an overview of the association results through a heatmap view where SNPs are plotted against a network of candidate genes, shows interactions between genes, integrates the association strengths of the genes to SNPs in the genome, and creates a tree view of structured genes to explore and identify functional relevant branches of the tree that

are associated with a genomic region. Although GenAMap was primarily developed for human diseases, it can be applied to plant AM as well.

PLINK software v. 1.07 (Purcell et al., 2007; http://pngu.mgh.harvard.edu/purcell/plink/) is an open source C/C++ GWAS tool set. With PLINK, large data sets comprising hundreds of thousands of SNPs and individuals can be readily manipulated and analyzed. PLINK offers five main characteristics. Data management is a simple interface for reordering, recording and filtering genotypic information. Summary statistics to determine the randomness of genotyping failure highlights the test of missingness on a simple haplotypic case-control test. Population stratification is measured on the basis of a genome average proportion of alleles sharing identical by state (IBS) between any two individuals. PLINK offers tools to cluster individuals into homogeneous subsets to identify potential outlier individuals causing genotyping or pedigree errors, and to incorporate this stratification in GWAS. Association analyses include CC, stratified analysis, TDT, QTDT, sib TDT and correction for multiple tests. Table 2 summarizes these and other software based on their analytical focus.

10. Future perspectives of AM

Large scale GWAS have already been carried out in plants and many more are in progress. The technological problem of efficiently genotyping 1 million or more SNPs has been solved, and the cost of genotyping continues to decline (Slatkin, 2008). With this increased resolution of LD patterns, the study of crop history will shift in focus from understanding the average history of populations to understanding the history of different genomic regions in depth. GWAS will not be limited to the identification of QTL but will also provide in depth understanding of the genomic changes that have shaped crop plants as a consequence of domestication and selection. Such information will translate into improved design of breeding populations and germplasm collections capturing adaptative variation.

Design and implementation of genotyping assays is no longer time-consuming or expensive. To fully exploit and benefit from the large amount of achievable genotyping data, care must be given to proper and powerful experimental design (Myles et al., 2009). Because LD mapping often involves a relatively large number of diverse accessions, phenotypic data collection with adequate replications across multiple years and locations can be challenging. Efficient field design, appropriate statistical methods and consideration for QTL × environmental interactions should be explored to increase the mapping power, particularly if field conditions are not homogeneous. Reducing errors associated with phenotypic measurements remains a priority.

One of the limitations of LD mapping is that it provides little insight into the mechanistic basis of LD detected, so that genomic localization and cloning of genes based on LD may not be always straightforward. This limitation occurs because strong LD is sometimes the result of a recent occurrence of LD (recent mutations) rather than a close physical linkage between two loci exhibiting LD (Gupta et al., 2005). As a consequence, we anticipate increased usage of nested AM, because it has the power to simultaneously capture information about the linkage of the markers and the degree of LD historically created. Linkage mapping and LD mapping are complementary and their successful combination has been demonstrated in plant systems (Brachi et al., 2010; Poland et al., 2011). As mentioned earlier, AM is one of

Software	Focus	Description	Website
STRUCTURE 2.3	Population structure	Compute a MCMC Bayesian analysis to estimate the proportion of the genome of an individual originating from the different inferred populations	http://pritch.bsd.uchicago.edu/software.html
BAPS 5.0	Population structure	Compute Bayesian analysis to estimate the proportion of the genome of an individual and assign individuals to genetic clusters by either considering them as immigrants or as descendents from immigrants	http://web.abo.fi/fak/mnf/mate/jc/software/baps.html
mStruct	Population structure	Detection of population structure in the presence of admixing and mutations from multi-locus genotype data. It is an admixture model which incorporates a mutation process on the observed genetic markers	http://www.cs.cmu.edu/~suyash/mstruct.html
LDheatmap	LD	R environment software for LD estimation (r^2) displayed as heatmap plots using SNPs	http://www.jstatsoft.org/v16/c03
LDhat 2.1	Recombination rates and LD	R environment software for LD estimation and identification of hotspots using a Bayesian reversible jump MCMC scheme for SNPs	http://www.stats.ox.ac.uk/mcvean/ldhat.html
MIDAS	LD	Compute LD heatmaps for D' and r^2 providing both inter loci and inter allelic LD variation	http://www.genes.org.uk/software/midas/
Arlequin 3.5	Genetic analysis and LD	Hierarchical analysis of genetic structure (AMOVA), LD for D' and r^2. Version 3.5 incorporates a R function to parse XML output files to produce publication quality graphics	http://cmpg.unibe.ch/software/arlequin35/

Software	Focus	Description	Website
Haploview 4.2	Haplotype analysis and LD	LD and haplotype block analysis, haplotype population frequency estimation, single SNP and haplotype association tests, permutation testing for association significance	http://www.broad.mit.edu/mpg/haploview/
GGT 2.0	Genetic analysis, LD and AM	Compute genetic distance based on Jaccard similarity, dendrograms are displayed using a Neighbour-Joining algorithm. Displays LD heatmaps and LD scatter plots for D' and r^2 and performs simple AM analysis	http://www.plantbreeding.wur.nl/UK/software_ggt.html
SVS 7	Stratification, LD and AM	Estimate stratification, LD, haplotypes blocks and multiple AM approaches for up to 1.8 million SNPs and 10,000 samples	http://www.goldenhelix.com
TASSEL	Stratification, LD and AM	SSR markers, GLM and MLM methods	http://www.maizegenetics.net
GenStat	Stratification, LD and AM	SSR markers, GLM and MLM-PCA methods	http://www.vsni.co.uk/
JMP genomics	Stratification, LD and structured AM	SNPs, CG and GWAS, analysis of common and rare variants	http://www.jmp.com/software/genomics
GenAMap	Stratification, LD and structured AM	SNPs, tree of functional branches, multiple visualization tools	http://cogito-b.ml.cmu.edu/genamap
PLINK	Stratification, LD and structured AM	SNPs, multiple AM approaches, IBD and IBS analyses	http://pngu.mgh.harvard.edu/purcell/plink/

Table 2. List of software used in LD and AM.

many applications of LD. With the increasing availability of molecular markers, it is now feasible to scan a genome to identify signatures of selection (both positive and balancing selection). This approach, known as population genomics, simultaneously studies thousands

of marker loci distributed across the genome to better understand the roles that evolutionary processes have played in the current pattern of genetic variation across populations (Luikart et al., 2003). Among different approaches reviewed by Oleksyk et al. (2010), LD can be used to Identify loci that have been targets of selection. Strong positive selection quickly increases the frequency of an advantageous allele, resulting in linked loci remaining in unusually strong LD with that allele in the phenomenon known as genetic hitchhiking. Since conditions vary from one locality to another and differ considerably between ecosystems (Oleksyk et al., 2010), it is expected that genomic differentiation occurred between populations. Patterns of contrasting LD between populations can assist in identifying adaptative genetic diversity for emerging global problems such as drought tolerance, UV radiation, heavy metal related genes, and ultimately, food security. Since climate change is likely to affect a wide range of species and habitats, LD could assist in the development of specific "adapted germplasm collections" suitable for cultivar development and conservation rather than collections capturing mostly neutral variation. Studies on the adaptation of natural populations to local ecosystems based on LD variations have already been reported (Li & Merilä, 2011).

Although GWAS have been successful in finding new causative alleles, usually tests for common variants are underpowered for detecting variants of lower frequency leaving a high proportion of undetected heritability. In human genetics, there is a growing interest in the role of rare variants in multifactorial disease etiology and there is an increasing body of evidence pointing to the role of rare variants in complex traits (Bansal et al., 2010). The frequency of any single rare or low-frequency variant is less than 5%, but collectively the number of rare variants is substantial. According to the multiple rare variant (MRV) hypothesis, there are many large effect rare variants in the population and cases of common inherited diseases have been the result of additive effects of a few of these moderate to high penetrance MRVs (Bodmer & Bonillna, 2008, as cited in Asimit & Zeggini, 2010). In the search for causal variants of type 1 diabetes (T1D), Nejentsev et al. (2009) identified four disease-associated rare variants in the *IFIH1* gene, which are protective of T1D. Involvement of rare variants in hypertension has also been shown (Ji et al., 2008). Despite their importance, rare variants have not been studied as extensively as common variants because of cost limitations in next generation sequencing technologies and the lack of an appropriate analytical toolbox to enable powerful rare variant association analysis (Asimit & Zeggini, 2010). With this in mind, several strategies for association studies involving rare variants have been proposed. The simplest approach is to test them individually using standard contingency table and regression methods such as those implemented in the genetic software PLINK (Purcell et al., 2007). This method, called "single-locus test" is highly problematic, given, for example, the poor power that such statistical tests have to detect small differences between diagnostic or phenotypic groups (Gorlov et al., 2008, as cited in Bansal et al., 2010). Other methods that overcome the power issues associated with testing rare variants individually include the collapsing strategy, methods based on summary statistics, multiple regression and data mining which are comprehensively reviewed by Bansal et al. (2010). Approaches involving direct sequencing have been tested by Li & Leal (2009). Since epigenetic factors are also likely to contribute to common complex traits, epigenome-wide association studies (EWASs) have been proposed to uncover another missing piece of heritability unexplained by common variants (Rakyan et al., 2011), specifically involving the study of variation in DNA methylation across the genome.

Future scenarios in plant AM will likely include a combination of studies involving common and rare variants to explain most of the phenotypic variation observed for agronomic and adaptative traits. The reader will have noticed the influence of human genetics in much of the discussion of LD mapping. Plant geneticists continue to follow human genetics research in order to improve QTL studies. However, plants offer advantages that cannot be afforded in humans such as population design and size, which promise to make plant GWAS a powerful tool. Overall, we anticipate witnessing advances in plant AM as a result of new approaches in human association studies in combination with the benefits of plant genetics that enable us to uncover and understand levels of plants genome complexity not seen before.

11. Acknowledgments

We are grateful to Andrzej Walichnowski for help with manuscript editing and Mike Shillinglaw for figures.

12. References

Abdurakhmonov, I. & Abdukarimov, A. (2008). Application of association mapping to understanding the genetic diversity of plant germplasm resources. *International Journal of Plant Genomics*, Vol. 2008, 1-18

Abdurakhmonov, I., Kohel, R., Yu, J., Pepper, A., Abdullaev, A., Kushanov, F., Salakhutdinov, I., Buriev, Z., Saha, S., Scheffler, B., Jenkins, J. & Abdukarimov. A. (2008). Molecular diversity and association mapping of fiber quality traits in exotic *G. Hirsutum* L. germplasm. *Genomics*, Vol. 92, No. 6, 478-487

Abdurakhmonov, I., Saha, S., Jenkins, J., Buriev, Z., Shermatov, S., Scheffler, B., Pepper, A., Yu, J., Kohel, R. & Abdukarimov A. (2009). Linkage disequilibrium based association mapping of fiber quality traits in *G. hirsutum* L. variety germplasm. *Genetica*, Vol. 136, No. 3, 401-417

Achleitner, A., Tinker, N., Zechner, E. & Buerstmayr, H. (2008). Genetic diversity among oat varieties of worldwide origin and associations of AFLP markers with quantitative traits. *Theoretical and Applied Genetics*, Vol. 117, No. 7, 1041-1053

Aranzana, M., Kim, S., Zhao, K., Bakker, E., Horton, M., Jacob, K., Lister, C., Molitor, J., Shindo, C., Tang, C., Toomajian, C., Traw, B., Zheng, H., Bergelson, J., Dean, C., Marjoram, P. & Nordborg M. (2005). Genome-wide association mapping in Arabidopsis identifies previously known flowering time and pathogen resistance genes. *PLoS Genetics*, Vol. 1, No. 5, e60

Ardlie, K., Kruglyak, L. & Seielstad, M. (2002). Patterns of linkage disequilibrium in the human genome. *Nature Review Genetics*, Vol. 3, No. 4, 299-309

Arnheim, N., Calabrese, P. & Nordborg, M. (2003). Hot and cold spots of recombination in the human genome: the reason we should find them and how this can be achieved. *The American Journal of Human Genetics*, Vol. 73, No. 1, 5-16

Asimit, J. & Zeggini, E. (2010). Rare variant association analysis methods for complex traits. *Annual Review of Genetics,* Vol. 44, 293-308

Balding, D. (2006). A tutorial on statistical methods for population association studies. *Nature Review Genetics*, Vol. 7, No. 10, 781-791

Bansal, V., Libiger, O., Torkamani, A. & Schork, N. (2010). Statistical analysis strategies for association studies involving rare variants. *Nature Review Genetics*, Vol. 11, No. 11, 773-785

Bar-Hen, A., Charcosset, A., Bourgoin, M. & Guiard, J. (1995). Relationship between genetic markers and morphological traits in a maize inbred line collection. *Euphytica*, Vol. 84, No. 2, 145-154

Beer, S., Siripoonwiwat, W., O'donoughue, L., Souza, E., Mathews, D. & Sorrels, M. (1997). Associations between molecular markers and quantitative traits in oat germplasm pool: can we infer linkages? *Journal of Agricultural Genomics*, Vol. 3

Brachi, B., Faure, N., Horton, M., Flahauw, E., Vazquez, A., Nordborg, M., Bergelson, J., Cuguen, J. & Roux, F. (2010). Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. *PLoS Genetics*, Vol. 6, No. 5, e1000940

Bradbury, P., Zhang, Z., Kroon, D., Casstevens, T., Ramdoss, Y. & Buckler, E. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, Vol. 23, No. 19, 2633-2635

Breseghello, F. & Sorrells, M. (2006). Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics*, Vol. 172, No. 2, 1165-1177

Cockram, J., White, J., Zuluaga, D., Smith, D., Comadran, J., Macaulay, M., Luo, Z., Kearsey, M., Werner, P., Harrap, D., Tapsell, C., Liu, H., Hedley, P., Stein, N., Schulte, D., Steuernagel, B., Marshall, D., Thomas, W., Ramsay, L., Mackay, I., Balding, D., AGOUEB Consortium., Waugh, R. & O'Sullivan, D. (2010). Genome-wide association mapping to candidate polymorphism resolution in the unsequenced barley genome. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 107, No. 50, 21611-21616

Dawson, E., Abecasis, G., Bumpstead, S., Chen, Y., Hunt, S., Beare, D., Pabial, J., Dibling, T., Tinsley, E., Kirby, S., Carter, D., Papaspyridonos, M., Livingstone, S., Ganske, R., Lõhmussaar, E., Zernant, J., Tõnisson, N., Remm, M., Mägi, R., Puurand, T., Vilo, J., Kurg, A., Rice, K., Deloukas, P., Mott, R., Metspalu, A., Bentley, D., Cardon, L. & Dunham, I. (2002). A first-generation linkage disequilibrium map of human chromosome 22. *Nature*, Vol. 418, No. 6897, 544-548

Du, F., Clutter, A. & Lohuis, M. (2007). Characterizing linkage disequilibrium in pig populations. *International Journal of Biological Sciences*, Vol. 3, No. 3, 166-178

Ersoz, E., Yu, J. & Buckler, E. (2007). Application of linkage disequilibrium and association mapping in crop plants, In: *Genomics-Assisted Crop Improvement*, R. Varshney, & R. Tuberosa, (Ed.), 97-119, Springer, ISBN 987-1-4020-6294-0, Dordrecht, the Netherlands

Evanno, G., Regnaut, S. & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, Vol. 14, No. 8, 2611-2620

Excoffier, L. & Heckel, G. (2006). Computer programs for population genetics data analysis: a survival guide. *Nature Review Genetics*, Vol. 7, No. 10, 745-758

Flint-Garcia, S., Thornsberry, J. & Bukler, E. (2003). Structure of linkage disequilibrium in plants. *Annual Review Plant Biology*, Vol. 54, 357-374

Garris, A., Tai, T., Coburn, J., Kresovich, S. & McCouch, S. (2005). Genetic structure and diversity in *Oryza sativa* L. *Genetics*, Vol. 169, No. 3, 1631-1638

Gaut, B. & Long, A. (2003). The lowdown of linkage disequilibrium. *The Plant Cell*, Vol. 15, No. 7, 1502-1506

Gupta, P., Rustgi, S. & Kulwal, P. (2005). Linkage disequilibrium and association studies in higher plants: present status and future prospects. *Plant Molecular Biology*, Vol. 57, No. 4, 461-485

Hardy, O. & Vekemans, X. (2002). SPAGeDi: a versatile computer program to analyze spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, Vol. 2, No. 4, 618-620

Hill, W. & Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics*, Vol. 38, No. 6, 226-231

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C., Zhu, C., Lu, T., Zhang, Z., Li, M., Fan, D., Guo, Y., Wang, A., Wang, L., Deng, L., Li, W., Lu, Y., Weng, Q., Liu, K., Huang, T., Zhou, T., Jing, Y., Li, W., Lin, Z., Buckler, E., Qian, Q., Zhang, Q., Li, J. & Han, B. (2010). Genome-wide studies of 14 agronomic traits in rice landraces. *Nature Genetics*, Vol. 42, No. 11, 961-967

Hubisz, M., Falush, D., Stephens, M. & Pritchard, J. (2009). Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, Vol. 9, No. 5, 1322-1332

Jennings, H. (1917). The numerical results of diverse systems of breeding, with respect to two pairs of characters, linked or independent, with special relation to the effects of linkage. *Genetics*, Vol. 2, No. 2, 97-154

Ji, W., Foo, J., O'Roak, B., Zhao, H., Larson, M., Simon, D., Newton-Cheh, C., State, M., Levy, D. & Lifton, R. (2008). Rare independent mutations in renal salt handling genes contribute to blood pressure variation. *Nature Genetics*, Vol. 40, No. 5, 592-599

Jung, M., Ching, A., Bhattramakki, D., Dolan, M., Tingey, S., Morgante, M. & Rafalski, A. (2004). Linkage disequilibrium and sequence diversity in a 500-kbp region around the *adh*1 locus in elite maize germplasm. *Theroretical and Applied Genetics*, Vol. 109, No. 4, 681-689

Krill, A., Kirst, M., Kochian, L., Buckler, E. & Hoekenga, O. (2010). Association and linkage analysis of aluminum tolerance genes in maize. *PLoS One*, Vol. 5, No. 4, e9958

Kump, K., Bradbury, P., Wisser, R., Buckler, E., Belcher, A., Oropeza-Rosas, M., Zwonitzer, J., Kresovich, S., McMullen, M., Ware, D., Balint-Kurti, P. & Holland, J. (2011). Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. *Nature Genetics*, Vol. 43, No. 2, 163-168

Lewontin, C. (1964). The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics*, Vol. 49, No. 1, 49-67

Li, B. & Leal, S. (2008). Methods for detecting associations with rare variants for common diseases: Application to analysis of sequence data. *The American Journal of Human Genetics*, Vol. 83, No. 3, 311-321

Li, M. & Merilä, J. (2011). Population differences in levels of linkage disequilibrium in the wild. *Molecular Ecology*, Vol. 20, No. 14, 2916-2928

Li, X., Wei, Y., Moore, K., Michaud, R., Viands, D., Hansen, J., Acharya, A. & Brummer, E. (2011a). Association mapping of biomass yield and stem composition in a tetraploid alfalfa breeding population. *The Plant Genome*, Vol. 4, No. 1, 24-35

Li, Y., Haseneyer, G., Schön, C., Ankerst, D., Korzun, V., Wilde, P. & Bauer, E. (2011b). High levels of nucleotide diversity and fast decline of linkage disequilibrium in rye (*Secale cereale* L.) genes involved in frost response. *BMC Plant Biology*, Vol. 11, No. 6, 1-14

Li, Y., Huang, Y., Bergelson, J., Nordborg, M. & Borevitz, J. (2010). Association mapping of local climate-sensitive quantitative trait loci in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 107, No. 49, 21119-21204

Liu, A. & Burke, J. (2006). Patterns of nucleotide diversity in wild and cultivated sunflowers. *Genetics*, Vol. 173, No. 1, 321-330

Liu, K. & Muse, S. (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics,* Vol. 21, No. 9, 2128-2129

Lu, Y., Zhang, S., Shah, T., Xie, C., Hao, Z., Li, X., Farkhari, M., Ribaut, J., Cao, M., Rong, T. & Xu, Y. (2010). Joint linkage-linkage disequilibrium mapping is a powerful approach to detecting quantitative trait loci underlying drought tolerance in maize. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 107, No. 45, 19585-19590

Luikart, G., England, P., Tallmon, D., Jordan, S. & Taberlet, P. (2003). The power and promise of population genomics: from genotyping to genome typing. *Nature Review Genetics*, Vol. 4, No. 12, 981-994

Mackay, I. & Powell, W. (2007). Methods for linkage disequilibrium mapping in crops. *Trends in Plant Science*, Vol. 12, No. 2, 57-63

Malysheva-Otto, L., Ganal, M. & Röder, M. (2006). Analysis of molecular diversity, population structure and linkage disequilibrium in a worldwide survey of cultivated barley germplasm (*Hordeum vulgare* L.). *BMC Genetics*, Vol. 7, No. 6, 1-14

Myles, S., Peiffer, J., Brown, P., Ersoz, E., Zhang, Z., Costich, D. & Buckler, E. (2009). Association mapping: critical considerations shift from genotyping to experimental design. *The Plant Cell*, Vol. 21, No. 8, 2194-2202

Myles, S., Boyko, A., Owens, C., Brown, P., Grassi, F., Aradhya, M., Prins, B., Reynolds, A., Chia, J., Ware, D., Bustamante, C. & Buckler, E. (2011). Genetic structure and domestication history of the grape. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 108, No. 9, 3530-3535

Nejentsev, S., Walker, N., Riches, D., Egholm, M. & Todd, J. (2009). Rare variants of *IFIH1*, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science*, Vol. 324, No. 5925, 387-389

Nemri, A., Atwell, S., Tarone, A., Huang, Y., Zhao, K., Studholme, D., Nordborg, M. & Jones, J. (2010). Genome-wide survey of Arabidopsis natural variation in downy mildew resistance using combined association and linkage mapping. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 107, No. 22, 10302-10307

Nielsen, D. & Zaykin, D. (2001). Association mapping: where we've been, where we're going. *Expert Review of Molecular Diagnostics*, Vol. 1, No. 3, 334-342

Noonan, J., Coop, G., Kudaravalli, S., Smith, D., Krause, J., Alessi, J., Chen, F., Platt, D., Pääbo, S., Pritchard, J. & Rubin, E. (2006). Sequencing and analysis of Neanderthal genomic DNA. *Science*, Vol. 314, No. 5802, 1113-1118

Nordborg, M. (2000). Linkage disequilibrium, gene trees and selfing: An ancestral recombination graph with partial self-fertilization. *Genetics*, Vol. 154, No. 2, 923-929

Nordborg, M., Borevitz, J., Bergelsom, J., Berry, C., Chory, J., Hagenbland, J., Kreitman, M., Maloof, J., Noyes, T., Oefner, P., Stahl, E. & Weigel, D. (2002). The extent of linkage disequilibrium in *Arabidopsis thaliana*. *Nature Genetics*, Vol. 30, No. 2, 190-193

Oleksyk, T., Smith, M. & O'Brien, S. (2010). Genome-wide scans for footprints of natural selection. *Philosophical Transactions of the Royal Society of London Biological Sciences*, Vol. 365, No. 1537, 185-205

Oraguzie, N., Wilcox, P., Rikkerink, H. & de Silva, H. (2007). Linkage disequilibrium, In: *Association mapping in plants*, N.C. Oraguzie, E.H.A. Rikkerink, S.E. Gardiner & H.N. de Silve, (Ed.), 11-39, Springer, ISBN-13 9780387358444, New York, USA

Palaisa, K., Morgante, M., Tingey, S. & Rafalski, A. (2003). Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. *The Plant Cell*, Vol. 15, No. 8, 1795-1806

Poland, J., Bradbury, P., Buckler, E. & Nelson, R. (2011). Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 108, No. 17, 6893-6898

Pritchard, J. & Rosenberg, N. (1999). Use of unlinked genetic markers to detect population stratification in association studies. *The American Journal of Human Genetics*, Vol. 65, No. 1, 220-228

Pritchard, J., Stephens, M., Rosenberg, N. & Donnelly, P. (2000). Association mapping in structured populations. *The American Journal of Human Genetics*, Vol. 67, No. 1, 170-181

Pritchard, J. (2001). Deconstructing maize population structure. *Nature Genetics*, Vol. 28, No. 3, 203-204

Price, A., Patterson, N., Plenge, R., Weinblatt, M., Shadick, N. & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, Vol. 38, No. 8, 904-909

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M., Bender, D., Maller, J., Sklar, P., de Bakker, P., Daly, M. & Sham, P. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, Vol. 81, No. 3, 559-575

Rakyan, V., Down, T., Balding, D. & Beck, S. (2011). Epigenome-wide association studies for common human diseases. *Nature Review Genetics*, Vol. 12, No. 8, 529-541

Raman, H., Stodart, B., Ryan, P., Delhaize, E., Emebiri, L., Raman, R., Coombes, N. & Milgate, A. (2010). Genome-wide association analysis of common wheat (*Triticum aestivum* L.) germplasm identifies multiple loci for aluminium resistance. *Genome*, Vol. 53, No. 11, 957-966

Reeves, P. & Richards, C. (2009). Accurate inference of subtle population structure (and other genetic discontinuities) using principal coordinates. *PLoS One*, Vol. 4, No. 1, e4269

Robbins, M., Sim, S., Yang, W., Deynze, A., van der Knaap, E., Joobeur, T. & Francis, D. (2011). Mapping and linkage disequilibrium analysis with a genome-wide collection of SNPs that detect polymorphism in cultivated tomato. *Journal of Experimental Botany*, Vol. 62, No. 6, 1831-1845

Rostoks, N., Ramsay, L., MacKenzie, K., Cardle, L., Bhat, P., Roose, M., Svensson, J., Stein, N., Varshney, R., Marshall, D., Graner, A., Close, T. & Waugh R. (2006). Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 103, No. 49, 18656-18661

Rousset, M., Bonnin, I., Remoué, C., Falque, M., Rhoné, B., Veyrieras, J., Madur, D., Murigneux, A., Balfourier, F., Le Gouis, J., Santoni, S. & Goldringer I. (2011). Deciphering the genetics of flowering time by an association study on candidate genes in bread wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics*, Vol. 123, No. 6, 907-926

Schmegner, C., Hoegel, J., Vogel, W. & Assum, G. (2005). Genetic variability in a genomic region with long-range linkage disequilibrium reveals traces of a bottleneck in the history of the European population. *Human Genetics*, Vol. 118, No. 2, 276-286

Schulze, T. & McMahon, F. (2002). Genetic association mapping at the crossroads: which test and why? Overview and practical guidelines. *American Journal of Medical Genetics*, Vol. 114, No. 1, 1-11

Slatkin, M. (2008). Linkage disequilibrium: Understanding the evolutionary past and mapping the medical future. *Nature Review Genetics*, Vol. 9, No. 6, 477-485

Smith, A., Thomas, D., Munro, H. & Abecasis, G. (2005). Sequence features in regions of weak and strong linkage disequilibrium. *Genome Research*, Vol. 15, No. 11, 1519-1534

Spielman, R., McGinnis, R. & Ewens, W. (1993). Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *The American Journal of Human Genetics,* Vol. 52, No. 3, 506-516

Stapley, J., Birkhead, T., Burke, T. & Slate, J. (2010). Pronounced inter- and intrachromosomal variation in linkage disequilibrium across the zebra finch genome. *Genome Research*, Vol. 20, No. 4, 496-502

Stich, B., Piepho, H., Schulz, B. & Melchinger, A. (2008). Multi-trait association mapping in sugar beet (*Beta vulgaris* L.). *Theoretical and Applied Genetics*, Vol. 117, No. 6, 947-954

Stich, B. & Melchinger, A. (2010). An introduction to association mapping in plants. *CAB Reviews: Perspectives in agriculture, Veterinary Science, Nutrition and Natural Resources*, Vol. 5, No. 39, 1-9

Tenaillon, M., Sawkins, M., Long, A., Gaut, R., Doebley, J. & Gaut, B. (2001). Patterns of DNA sequence polymosphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 98, No. 16, 9161-9166

Teo, Y., Fry, A., Bhattacharya, K., Small, K., Kwiatkowski, D. & Clark, T. (2009). Genome-wide comparisons of variation in linkage disequilibrium. *Genome Research*, Vol. 19, No. 10, 1849-1860

Thornsberry, J., Goodman, M., Doebley, J., Kresovich, S., Nielsen, D. & Buckler, E. (2001). *Dwarf8* polymorphisms associate with variation in flowering time. *Nature Genetics*, Vol. 28, No. 3, 286-289

Tian, F., Bradbury, P., Brown, P., Hung, H., Sun, Q., Flint-Garcia, S., Rocheford, T., McMullen, M., Holland, J. & Buckler, E. (2011). Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nature Genetics*, Vol. 43, No. 2, 159-162

Virk, P., Ford-Lloyd, B., Jackson, M., Pooni, H., Clemeno, T. & Newbury, H. (1996). Predicting quantitative variation within rice germplasm using molecular markers. *Heredity*, Vol. 76. No. 3, 296-304

Von Zitzewitz, J., Cuesta-Marcos, A., Condon, F., Castro, A., Chao, S., Corey, A., Filichkin, T., Fisk, S., Gutierrez, L., Haggard, K., Karsai, I., Muehlbauer, G., Smith, K., Veisz, O. & Hayes, P. (2011). The genetics of winterhardiness in barley: perspectives from genome-wide association mapping. *The Plant Genome*, Vol. 4, No. 1, 76-91

Wang, J., McClean, P., Lee, R., Goos, R. & Helms, T. (2008). Association mapping of iron deficiency chlorosis loci in soybean (*Glycine max* L. Merr.) advanced breeding lines. *Theoretical and Applied Genetics*, Vol. 116, No. 6, 777-787

Wang, M., Sukumaran, S., Barkley, N., Chen, Z., Chen, C., Guo, B., Pittman, R., Stalker, H., Holbrook, C., Pederson, G. & Yu, J. (2011). Population structure and marker-trait association analysis of the US peanut (*Arachis hypogaea* L.) mini-core collection. *Theoretical and Applied Genetics*, Vol. 123, No. 8, 1307-1317

Weber, A., Zhao, Q., McMullen, M. & Doebley, J. (2009). Using association mapping in teosinte to investigate the function of maize selection-candidate genes. *PLoS One*, Vol. 4, No. 12, e8227

Wen, W., Mei, H., Feng, F., Yu, S., Huang, Z., Wu, J., Chen, L., Xu, X. & Luo, L. (2009). Population structure and association mapping on chromosome 7 using a diverse panel of Chinese germplasm of rice (*Oryza sativa* L.). *Theoretical and Applied Genetics*, Vol. 119, No. 3, 459-470

Wright, S. & Gaut, B. (2005). Molecular population genetics and the search for adaptative evolution in plants. *Molecular Biology and Evolution*, Vol. 22, No. 3, 506-519

Würschum, T., Maurer, H., Kraft, T., Janssen, G., Nilsson, C. & Reif, J. (2011). Genome-wide association mapping of agronomic traits in sugar beet. *Theoretical and Applied Genetics*, Vol. 123, No. 7, 1121-1131

Yan, J., Shan, T., Warburton, M., Buckler, E., McMullen, M. & Crouch, J. (2009). Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. *Plos One*, Vol. 4, No. 12, e8451

Yin, T., DiFazio, S., Gunter, L., Jawdy, S., Boerjan, W. & Tuskan, G. (2004). Genetic and physical mapping of *Melampsora* rust resistance genes in *Populus* and characterization of linkage disequilibrium and flanking genomic sequence. *New Phytologist*, Vol. 164, No. 1, 95-105

Yu, J. & Buckler, E. (2006). Genetic association mapping and genome organization of maize. *Current Opinions in Biotechnology*, Vol. 17, No. 2, 155-160

Yu, J., Pressoir, G., Briggs, W., Vroh Bi, I., Yamasaki, M., Doebley, J., McMullen, M., Gaut, B., Nielsen, D., Holland, J., Kresovich, S. & Buckler, E. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, Vol. 38, No. 2, 203-208

Yu, L., Lorenz, A., Rutkoski, J., Singh, R., Bhavani, S., Huerta-Espino, J. & Sorrells, M. (2011). Association mapping and gene-gene interaction for stem rust resistance in CIMMYT spring wheat germplasm. *Theoretical and Applied Genetics*, Vol. 123, No. 8, 1257-1268

Zhang, D., Bai, G., Zhu, C., Yu, J. & Carver, B. (2010). Genetic diversity, population structure, and linkage disequilibrium in U.S. elite winter wheat. *The Plant Genome*, Vol. 3, No. 2, 117-127

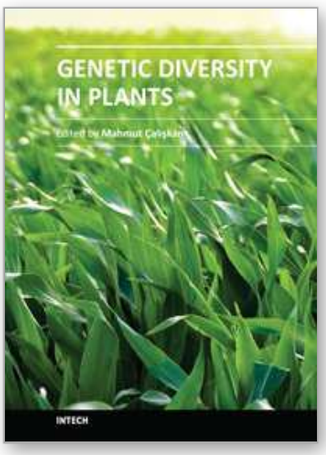
Zhao, K., Aranzana, M. & Kim, S. (2007a). An Arabidopsis example of association mapping in structured samples. *PLoS Genetics*, Vol. 3, No. 1, e4

Zhao, J., Paulo, M., Jamar, D., Lou, P., van Eeuwijk, F., Bonnema, G., Vreugdenhil, D. & Koornneef, M. (2007b). Association mapping of leaf traits, flowering time, and phytate content in *Brassica rapa*. *Genome*, Vol. 50, No. 10, 963-973

Zhu, Q., Zheng, X., Luo, J., Gaut, B. & Ge, S. (2007). Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: severe bottleneck during domestication of rice. *Molecular Biology and Evolution*, Vol. 24, No. 3, 875-888

Zhu, C., Gore, M., Buckler, E. & Yu, J. (2008). Status and prospects of association mapping in plants. *The Plant Genome*, Vol. 1, No. 1, 5-20

Zou, J., Jiang, C., Cao, Z., Li, R., Long, Y., Chen, S. & Meng, J. (2010). Association mapping of seed oil content in *Brassica napus* and comparison with quantitative trait loci identified from linkage mapping. *Genome*, Vol. 53, No. 11, 908-916



Genetic Diversity in Plants

Edited by Prof. Mahmut Caliskan

ISBN 978-953-51-0185-7
Hard cover, 498 pages
Publisher InTech
Published online 14, March, 2012
Published in print edition March, 2012

Genetic diversity is of fundamental importance in the continuity of a species as it provides the necessary adaptation to the prevailing biotic and abiotic environmental conditions, and enables change in the genetic composition to cope with changes in the environment. Genetic Diversity in Plants presents chapters revealing the magnitude of genetic variation existing in plant populations. The increasing availability of PCR-based molecular markers allows the detailed analyses and evaluation of genetic diversity in plants and also, the detection of genes influencing economically important traits. The purpose of the book is to provide a glimpse into the dynamic process of genetic variation by presenting the thoughts of scientists who are engaged in the generation of new ideas and techniques employed for the assessment of genetic diversity, often from very different perspectives. The book should prove useful to students, researchers, and experts in the area of conservation biology, genetic diversity, and molecular biology.

INTECH
open science | open minds

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

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

IntechOpen