Journal of Biological and Environmental Engineering
Vol. 2, No. 1, 2017, pp. 1-9
http://www.aiscience.org/journal/jbee

Association Mapping, Principles and Techniques

Ujjawal Kumar Singh Kushwaha*, Vikas Mangal, Anil Kumar Bairwa, Sneha Adhikari, Tabassum Ahmed, Pallavi Bhat, Ankit Yadav, Narendra Dhaka, Deepak Raj Prajapati, Amit Kumar Gaur, Ranjana Tamta, Indra Deo, Narendra Kumar Singh

Department of Genetics and Plant Breeding, College of Agriculture, Govind Ballabh Pant University of Agriculture and Technology, Pantnagar, Uttrakhand, India

Abstract

Association mapping is a powerful tool for the dissection of complex agronomic traits and for the identification of alleles. It is a very efficient and effective method for confirming candidate genes or for identifying new genes. Association mapping is a useful alternative to standard QTL mapping approaches which involves the correlation of molecular polymorphisms with phenotypic variation in a diverse assemblage of individuals. The comparatively high-resolution provided by association mapping is based on the structure of linkage disequilibrium (LD) across the genome. Many factors adversly affects association mapping, including population structure, small sample size, and low frequency of specific alleles that may increase the detection of a false positive associations. Association mapping offers great potential to enhance crop genetic improvement. Still extensive research is needed to understand the expansive application of association mapping. This review describes association mapping in detail along with its principles, techniques and application.

Keywords

Association Mapping, Linkage Disequilibrium, Population Size, Crop Improvement, Alleles

Received: January 30, 2017 / Accepted: April 8, 2017 / Published online: July 27, 2017

1. Background

Quantitative trait locus (QTL) mapping is a powerful and well-established tool for studying the genetic basis of complex quantitative traits in plants and animals (Yan et al., 2011). The recent advances in the development of unbiased association mapping approaches with its successful applications in dissecting a number of simple to complex traits in many crop species demonstrate powerful gene tagging tool for crops in the plant genomics era of 21st century.

2. Association Mapping

Association mapping is a high-resolution method for mapping quantitative trait loci based on principle of linkage disequilibrium that holds a great promise for the dissection of complex genetic traits (Buckler, 2002). It is a powerful tool for the dissection of complex agronomic traits and for the identification of alleles that can contribute to the enhancement of a target trait. The power of association studies is determined by the size of the experimental population, the magnitude of the target allele effect, the density of markers used, and the rate of LD decay between marker and target allele as well as errors in phenotyping and genotyping data and the desired resultant statistical significance level (Gordon and Finch, 2005).

Association mapping is a very efficient and effective method

* Corresponding author
E-mail address: kushujjwal@gmail.com (U. K. S. Kushwaha)

for confirming candidate genes or for identifying new genes (Altshuler et al., 2008; Hunter and Crawford, 2008). It is now being increasingly used in a wide range of plants (Rafalski, 2010), where it appears to be more powerful than in humans or animals (Zhu et al., 2008). Though association mapping is widely used, it has a lower power to detect rare alleles in a population, even those with large effects, than linkage mapping (Visscher, 2008). Association mapping is a useful alternative to standard QTL mapping approaches which involves the correlation of molecular polymorphisms with phenotypic variation in a diverse assemblage of individuals. The comparatively high-resolution provided by association mapping is dependent upon the structure of linkage disequilibrium (LD) across the genome. Association studies can be divided into two broad categories:

(i) Candidate gene association mapping

Variation in a gene of interest is tested and correlated with the phenotypic trait of interest.

(ii) Genome Wide Association mapping

Here, genetic variation is explored within the whole genome, aiming to find signals of association with the complex trait. GWA mapping is a promising method to identify novel loci involved in complex phenotypic traits. However, GWA mapping should not be regarded as a replacement of traditional QTL mapping. In fact, GWA mapping and QTL mapping have complementary advantages and disadvantages, which can lead to a better understanding of causal genetic polymorphism when these approaches are combined (Yu et al 2006).

Table 1. Comparison of family based (QTL) and population based (association mapping) methods that aim to unravel the genetic basis of complex trait in plants.

	QTL mapping	Candidate gene association mapping	Genome wide association mapping
Main advantage	No population structure effects, identification of rare alleles, few genetic markers required	Allow fine mapping, relatively low costs	Allows untargeted fine mapping (blind approach), detection of common alleles
Main disadvantage	Limited genetic diversity, not always possible to create crosses, can not distinguish between pleiortopic and physically close genes	Detailed functional knowledge of trait is required, no novel trait will be found	Confounding effect due to population structure, will miss rare and weak effect alleles
General requirements	Small original population size, low number of genetic markers, many replicates needed generated mapping materials e.g F2 population, (Al)RILs, MAGIC Lines, NILs, HIFs etc.	Large population size, small number of genetic markers, the bigger the population size, the less replicates needed, Prior genetic and biochemical knowledge on trait of interest, Prior knowledge on LD, nucleotide polymorphism, breeding system and population structure	Large population size, many genetic markers, The bigger the population size, the fewer replicates needed, Prior knowledge on LD, nucleotide polymorphism, breeding system and population structure

Schematic comparison of the main characteristics of different mapping strategies (following Darvasi and Shifman 2005)

	Linkage analysis	Admixture mapping	Joint linkage and LD mapping (and inbred-by-outbred cross)	Nested association mapping	Association mapping
Allele richness	Low	Low	Intermediate	High	High
Inference from markers in identity-by-state to quantitative trait nucleotides in IBD	Low	Low to intermediate	Intermediate	High	High
No. of SNPs required for whole-genome scan	Low	Low	Intermediate to high	Low (only high for founders)	High
Efficiency in using sequence information	Low	Low	Intermediate	High	Intermediate
Mapping resolution	Poor	Intermediate	Intermediate	Good	Good
Designed mapping population	Yes or no	Yes or No	Mostly no	Yes	No
Sensitivity to genetic heterogeneity	Low	Moderate	High	Low	High
Repeated phenotyping	Possible	Possible	Possible	Yes	Possible
Statistical power	Low to intermediate[a]	High[b]	Intermediate	High	High

[a] With designed mapping populations such as F_2, BC, or RIL, the power of linkage analysis is generally higher in plants than in humans.

[b] Power diminishes to zero with equal allele frequencies in the ancestral population (Darvasi and Shifman 2005).

2.1. Candidate Gene Strategy

The candidate gene method of association analysis is a hypothesis-driven approach for complex trait dissection that aims to identify the most important alleles. It involves genotyping or resequencing the genes considered to have a high probability of association with the phenotype(s) of interest within the germplasm being tested. There are a number of different approaches for implementing this strategy depending on the method used to identify the candidate gene and the level of confidence, the researcher has the belief that a given gene is important for the target trait. Earlier, it was common to sequence the gene of interest as fully as possible across a limited number of diverse lines (typically 24 to 48) to identify possible causal polymorphisms, such as SNPs causing amino acid changes or translated regions. The selected polymorphisms were then screened across a larger germplasm collection (of hundreds or thousands of genotypes) using inexpensive PCR-based SNP and/or indel genotyping assays (rather than sequencing) to confirm the associations between genotype and phenotype. In another method, the partial or entire gene is sequenced in all individuals of a germplasm panel (of several hundred genotypes) to identify significant associations, either with the causal polymorphism(s) or a polymorphism that is within LD distance to a causal polymorphism.

Although this is a more expensive approach, it may identify rare polymorphisms that can be missed by the first strategy. Determining which method to use has generally been based on the level of funding and the amount of time available for each study. However, resequencing of the entire gene has the added advantage that it can directly identify the best haplotype for each target breeding purpose.

2.2. Genome Wide Association Mapping

Genome Wide Association Studies (GWAS) have been recently used to dissect complex quantitative traits and to identify candidate genes affecting phenotype variation of polygenic traits. With the recent development of high-throughput genotyping technologies, genetic variation in many model organisms such as mice, arabidopsis, and maize is being discovered on a genome wide scale (Flint-Garcia et al 2005). Genome wide association mapping in model organisms has great potential to identify risk factors for complex traits related to human diseases.

Quantitative trait locus (QTL) mapping (Brotman et al 2011, Dobon et al 2011 and Balasubramanian et al., 2009) and association mapping are the most commonly used tools for dissecting the genetic basis of phenotypic trait variation. In QTL mapping only a limited number of recombination events that have occurred within families and pedigrees can be studied, whereas with association mapping the recombination events that have accumulated over thousands of generations can be exploited (Zhu et al., 2008). Since the 1980s, QTL mapping has been used most frequently, but association mapping is a promising alternative method for dissecting complex traits (Chan et al., 2010; Chan et al., 2011). Increased mapping resolution, reduced research time and larger allele numbers have been put forward as main advantages over traditional QTL mapping (Chan et al., 2010, Chan et al., 2011).

3. Factors Affecting Association Mapping

Many factors adversly affects association mapping, including population structure, small sample size, and low frequency of specific alleles that may increase the detection of a false positive associations.

Population structure significantly influences the efficiency of association mapping. The presence of population stratification and an unequal distribution of alleles facilitate mapping and identification of the underlying causes of quantitative trait variation in plants. Subgroups can result in non-functional, spurious associations. Highly significant LD between polymorphisms on different chromosomes may produce associations between a marker and a phenotype, even though the marker is not physically linked to the locus responsible for the phenotypic variation (Pritchard and Rosenberg, 1999). The complex breeding history of many important crops and the limited gene flow in most wild plants have created complex stratification within the germplasm, which complicates association studies. Association tests that do not attempt to account for the effects of population structure that must be viewed with skepticism.

Besides physical distance on the chromosome, many factors affect the breakdown of LD, including genetic drift, natural and artificial selection, mating system, and admixture of different populations (Flint-Garcia et al., 2003; Gaut and Long 2003; Yu and Buckler, 2006).

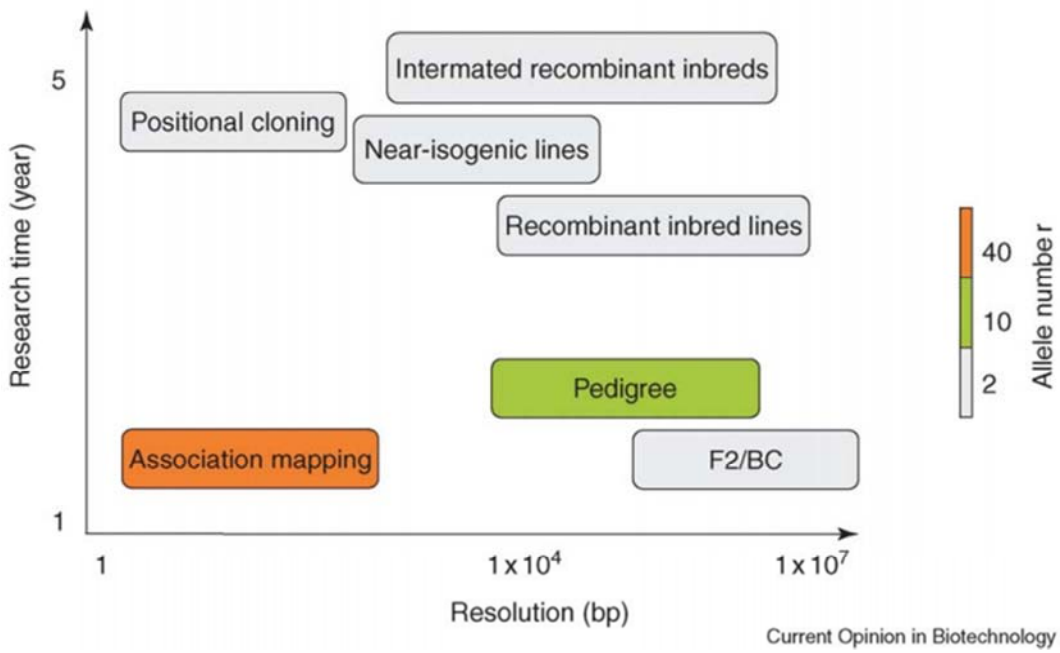

Figure 1. Schematic comparison of various methods for identifying nucleotide polymorphism trait association in terms of resolution, research time and allele number. BC, backcross traits (Yu and Buckler, 2006).

Population structure can cause some allele frequencies to differ significantly between subpopulations, which can create unexpected LD between unlinked loci across the genome (Ersoz et al., 2009). Neutral markers can be used to estimate population structure using traditional finger-printing and diversity analyses. Several statistical methods have been used to control the effect of population structure in association analyses including genomic control (Devlin and Roeder 1999; Mackay and Powell, 2007), structured association (Pritchard et al., 2000), principal components analysis (PCA) (Patterson et al., 2006, Price et al., 2006), non metricmulti dimensional scaling (nMDS) (Zhu and Yu, 2009), and the unified mixed-model approach (Flint-Garcia et al.,2005). A two stage dimension determination approach using both PCA and nMDS has been demonstrated to be the best approach to capture the major structure of association panels to maximize the rejection of false positives while maximizing the statistical power to identify real associations (Zhu and Yu, 2009).

Precise phenotyping is another key constraint for any marker–trait association analysis (Rafalski, 2010). In animal systems it is very difficult to obtain replicated phenotypic measurements for each genotype, but in plants it is relative easy to gener-ate pure breeding or homogenous lines for phenotyping in replicated trials across multiple environments and seasons.

Once markers have been identified that have been shown to be tightly and robustly linked to the target trait, they provide several magnitudes of return on investment through increased speed and cost efficiency of breeding programs.

4. General Procedure of Association Mapping

The exact details of the procedure depend on the chosen study design and the population structure (Singh and Singh 2016). The general procedure for genome-wide association mapping in plants is briefly outlined here based on Abdurakhmonov and Abdukarimov (2008).

4.1. Association Mapping Population

A large random sample from a natural population, a collection of breeding lines including cultivars, or a population derived from multi parent crosses of the concerned species use for association mapping. The sample should include as much genetic diversity present in the population collection as is practically feasible. This sample constitutes the association mapping population, association mapping panel, or association panel.

4.2. Phenotyping

The selected sample grows in field and morphologically evaluates the various traits of interest; this is called phenotyping. Phenotyping should preferably based on replicated trials conducted over locations and years to minimize environmental effects. The trials should conduct using a suitable experimental design like randomized block design, augmented design, nested design, etc. A precise and reliable phenotyping is critical to any mapping effort.

4.3. Genotyping for Population Structure Analysis

The sample then, goes for genotyping, i.e., tested with a set of molecular markers (preferably SSR markers) that are evenly distributed over the entire genome of the species. These markers should are unlinked, i.e., is located more than 40 cM apart in the genome (Pritchard et al., 2000a, b).

4.4. Structure and Kinship Analysis

The marker data then, analyze to detect and estimate the population structure of the sample using the STRUCTURE program and the extent of kinship among the individuals of the sample using the TASSEL program.

4.5. Genotyping for LD Analysis

The sample also genotyped with a sufficiently large number of molecular markers that cover the entire genome as densely as is feasible so that LD between markers and the loci of interest can be detected. The pattern of LD in the concerned genomic regions of the species and the extent of LD observed among different populations of the species would determine the number of markers required for adequate coverage of the whole genome. SSR and SNP marker systems are the most widely used for this purpose.

4.6. AM and LD Analyses

A model-based analysis of relatedness between the phenotype and the genotype data done to detect and quantify LD between the markers and the genes/QTLs governing the traits of interest. The estimates of population structure and kinship use as covariates in the model to minimize false associations between the markers and the genes/QTLs of interest. Since these analyses are computationally intensive, suitable computer programs use for their implementation.

5. Statistical Approaches Uses for Association Mapping

Recent developments in statistical methodologies make it possible to properly interpret the results of association tests. Pritchard et al., (2000) have developed an approach that incorporates estimates of population structure directly into the association test statistic. The essential idea of this method is to decompose a sample drawn from a mixed population into several unstructured subpopulations and test the association in the homogeneous subpopulations. The methods have been applied to association analyses in humans (Rosenberg et al., 2002;Cardon and Bell, 2001) and crop plants, with modified test statistics being used to deal with quantitative traits (Thornsberryet al., 2001).

LD between a single marker and a QTL can be measured by regression analysis, where the data on the trait is regressed on the individual marker genotypes, so that significant regressions will identify the markers associated with the phenotype (Remington et al., 2001).

Nowadays several software use to assess the association of marker loci with traits. The most commonly used statistics include logistic regression with the possibility of structured associations implemented in TASSEL General Linear Model (Yu and Buckler, 2006, TASSEL: http://www.maizegenetics.net), a multiple regression model combined with the estimates for the false discovery rate suggested by Kraakman et al., (2006), and an unified mixed-model approach described by Yu et al., (2006) and implemented in TASSEL Mixed Linear Model or in SAS v9.1.2 (Ehrenreichet al., 2007).

6. Advantage of Association Mapping

Association mapping is a valuable tool for the detection of novel genes or QTLs of important agronomic characteristics. The extensive application of this approach in crop plants is expected in the long term as a result of establishment of the novel high-throughput genotyping and sequencing technologies (Mackay and Powell, 2007;Oraguzieet al., 2007).

Gene-based markers are more accurate than linked markers for the prediction of phenotype, since the marker–trait association do not lost during segregation in the course of recurrent breeding selection cycles. Results from association analysis can be used to predict the best haplotype across one or multiple genes for optimum expression of the target trait.

Genome-wide association studies are currently exploited for mapping of disease genes in human genetics (The Wellcome Trust Case Control Consortium, 2007). In crop plants, the potential of exploiting LD to detect marker-trait associations was recently investigated for maize (Yu and Buckler 2006), wheat (Ravel et al., 2006; Tommasini et al., 2007), barley (Kraakman et al., 2004; Kraakman et al., 2006; Maly-sheva-Otto and Röder 2006; Rostoks et al., 2006), sorghum (Hamblin et al., 2004), ryegrass (Xing et al., 2007), soybean (Hyten et al., 2007) and rice (Garriset al., 2003).

Association studies based on correlations between alleles at different sites or LD can provide high resolution for the identification of genes that contribute to phenotypic variation in natural populations. This approach has a potential to identify a single polymorphism within a gene that is responsible for the difference in phenotype. In addition, many plant species have high levels of diversity for which association approaches are well suited to evaluate the numerous alleles available. LD plays a central role in association analysis. The distance over

which LD persists will determine the number and density of markers, and experimental design needed to perform an association analysis.

7. Linkage Disequilibrium

Linkage disequilibrium (LD) refers to the non-random association of alleles between genetic loci. The term was originally defined in relation to the population of alleles that reside on the same chromosome. Although LD is a population-based phenomenon, it is generally observed that there tends to be a higher LD between alleles that are located more closely together. Thus, the random association between alleles might be reduced by linkage thereby creating the so called disequilibrium. Many genetic and non-genetic factors, including recombination, drift, selection, mating pattern, and admixture, affect the structure of LD. The key to association mapping is the LD between functional loci and markers that are physically linked. Thus linkage disequilibrium is an important factor in association mapping. Several statistical parameters can be used to estimate the extent of LD (Hedrick, 1987), most commonly r^2, which estimates the correlation between allelic states of two given polymorphic loci. Linkage disequilibrium can be greatly over estimate when sample sizes smaller than 50 individuals are used (Yan et al., 2009).

In general, genetic linkage mapping studies identify linkage between a marker and the more distant functional DNA sequence by creating biparental mapping populations that have experienced only a few generations of recombination.

Studies have shown that LD levels vary both within and between species (Flint et al., 2003). For example, LD extends less than 1000 bp (Tenaillon et al., 2001) for maize landraces and roughly 2000 bp for diverse maize inbred lines, but can be as high as 100 kb for commercial elite inbred lines (Ching et al., 2001). LD decay can also vary considerably from locus to locus (Yan et al 2009). This may be due to the great variation in recombination rates along the chromosomes, including a low recombination rate in centromeric regions and a high recombination rate within genic regions due to retrotransposon insertions (Dooner and He., 2008). For example, significant LD was observed up to 4 kb for theY1 locus (encoding phytonene synthase), but was seen at only 1 kb forPSY2 (a putative phytonene synthase) in the same maize population (Palaisa et al., 2003).

8. Differences Between Linkage Analysis and Association Mapping

The most commonly used tools for dissecting genetic architecture of complex traits are linkage analysis and association mapping (Doerge,2002). Linkage analysis exploits the shared inheritance of functional polymorphisms and adjacent markers within families or pedigrees of known ancestry. Linkage analysis in plants has been typically conducted with experimental populations that are derived from a biparental cross. Although based on the same fundamental principles of genetic recombination as linkage analysis, association mapping examines the shared inheritance for a collection of individuals often with unobserved ancestry. As the unobserved ancestry can extend thousands of generations, the shared inheritance will only persist for adjacent loci after these many generations of recombination. Essentially, association mapping exploits historical and evolutionary recombination at the population level (Thornsberry et al., 2001, Remington et al., 2001). By exploring deeper population genealogy rather than family pedigree, association mapping offers three advantages over linkage analysis:

1 Much higher mapping resolution;

2 Greater allele number and broader reference population; and

3 Less research time in establishing an association (Buckler and Thornsberry, 2002; Flint et al., 2003).

Linkage analysis and association mapping, however, are complimentary to each other in terms of providing prior knowledge, cross-validation, and statistical power. As complementary approaches, linkage analysis often identifies broad chromosome regions of interest with relatively low marker coverage, while association mapping offers high resolution with either prior information on candidate genes or a genome scan with very high marker coverage (Hirschhorn and Daly, 2005). An integrated mapping strategy would combine the advantages of the two approaches to improve mapping resolution without requiring excessively dense marker maps. The possibility of developing such an integrated mapping strategy exists for the model species maize (*Zea mays* L.), because of the availability of a highly diverse collection of germplasm and the feasibility of creating segregating progenies and immortal genotypes through self-fertilization (Flint-Garcia et al., 2005).

9. Application of Association Mapping

Linkage disequilibrium can be used for a variety of purposes in crop plant genomics research. One of the major uses of LD in plants would be to study marker-trait association followed by marker-assisted selection (MAS). Another important application is its use in the studies of population genetics and

genetic diversity in natural populations and germplasm collections and in crop improvement programmes. Marker-trait association in crop plants is generally conducted through linkage analysis, utilizing methods like t-test, simple regression analysis and QTL interval mapping (Hackett, 2002). The limitations of linkage analysis approach imposed by the availability of mapping populations have largely been overcome in LD-based association mapping, which can be applied to germplasm bank collections, synthetic populations, and elite germplasm(Mackay 2001; Hackett 2002). Genetic association mapping or linkage disequilibrium mapping is a method that relies on linkage disequilibrium to study the relationship between phenotypic variation and genetic polymorphisms (Breseghello and Sorrells,2006).

The use of LD for mapping of QTLs for a quantitative trait is more challenging, but is also more rewarding, because it allows more precise locating of the position of a QTL controlling the trait of interest. When comparing linkage analysis and LD mapping for QTL detection, it is revealed that linkage mapping is more useful for genome-wide scan for QTLs, while LD mapping gives more precise location of an individual QTL. One may therefore like to use linkage analysis for preliminary location of QTLs and then use LD for more precise location (Mackay, 2001; Glazier et al., 2002).

Genetic association mapping is a new approach which takes into account thousands of polymorphisms to evaluate for QTL effect and is more efficient as compared to linkage analysis because it does not require generation of segregating populations/large numbers of progeny (Oraguzie et al., 2007). However, association mapping is only capable of identifying phenotypic effects of alleles with reasonably high frequency in the population under investigation. Rare alleles usually cannot be evaluated because of lack of power (not enough individuals carrying this allele). So, for such alleles classical biparental mapping can be more appropriate.

10. Summaries

Association mapping offers great potential to enhance crop genetic improvement. This is strengthened by the use of high throughput and cost effective next generation sequencing techniques that will enable GWA studies to become a popular and routine approach. However, association mapping remains complementary as replacement for linkage mapping and other gene identification and validation techniques. Moreover, the contrast between the large numbers of variants with small effects identified by GWA studies versus the small number of genomic regions with large effects identified by linkage mapping remains a challenge to our current understanding of the genetic architecture of complex traits. Although, for practical applications, the integration of linkage mapping and association mapping approaches offers substantial opportunity to resolve the individual constraints of each approach while synergizing their respective strengths. Population structure remains a big limitation for association studies that requires careful choice of germplasm and the development of advanced statistical approaches.

References

[1] Altshuler, D., M.J. Daly, and E.S. Lander. 2008. Genetic mapping in human disease. Science 322(5903):881–888.

[2] Balasubramanian, S. et al. 2009. QTL mapping in new *Arabidopsis thaliana* advanced intercross-recombinant inbred lines. PLoS ONE4, e4318.

[3] Breseghello, F. and M.S. Sorrells (2006) Association mapping of kernel size and milling quality in wheat (TriticumaestivumL.) cultivars. Genetics 172:1165-1177.

[4] Brotman, Y. et al. 2011. Identification of enzymatic and regulatory genes of plant metabolism through QTL analysis inArabidopsis. J. Plant Physiol. 168, 1387–1394.

[5] Buckler, E.S. and JM Thornsberry 2002. Plant molecular diversity and applications to genomics. CurrOpin Plant Biol 5: 107–111.

[6] Cardon, L.R., and J.I. Bell 2001. Association study designs for com-plex diseases. Nat Rev Genet 2:91-99.

[7] Chan, E.K.F. et al. 2011. Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in Arabidopsis thaliana. PLoS Biol.9, e1001125.

[8] Chan, E.K.F. et al. 2010. Understanding the evolution of defense metabolites inArabidopsisthalianausing genome-wide association mapping.Genetics185, 991–1007

[9] Ching A., K.S., Caldwell, M. Jung, M. Dolan, O.S. Smith, S. Tingey, M. Morgante, A.J., Rafalski 2001. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines.BMC Genet, 3:19.

[10] Devlin, B., and K. Roeder. 1999. Genomic control for association studies. Biometrics 55:997–1004.

[11] Dobon, A. et al. 2011. Quantitative genetic analysis of salicylic acid perception inArabidopsis. Planta234, 671–684.

[12] Doerge, R.W. 2002. Mapping and analysis of quantitative trait loci in experimental populations. Nat. Rev. Genet. 3:43–52.

[13] Dooner, H.K., and L. He 2008. Maize genome structure variation: Inter-play between retrotransposon polymorphisms and genic recombina-tion. Plant Cell 20:249–258.

[14] Ehrenreich I.M., Stafford P.A. and Purugganan M.D. 2007. The ge-netic architecture of shoot branching in Arabidopsis thaliana: A comparative assessment of candidate gene asso-ciations vs. quantitative trait locus mapping. Genetics 176:1223-1236.

[15] Ersoz, E.S., J. Yu, and E.S. Buckler, IV. 2009. Applications of linkage dis-equilibrium and association mapping in maize. p. 173–195. InA. L. Kriz and B. A. Larkins (ed.) Molecular genetic approaches to maizeimprovement. Springer-Verlag, New York, NY.

[16] Flint-Garcia S. A., J. M. Thornsberry, ES Buckler 2003. Structure of linkage disequilibrium in plants. Annu Rev Plant Biol,54:357-374.

[17] Flint-Garcia, S. A., A. C. Thuillet, J. Yu, G. Pressoir, S. M. Romero, S. E. Mitchell, J. Doebley, S. Kresovich, M. M. Goodman, and E. S. Buckler. 2005. Maize association population: A high-resolution platform for quantitative trait locus dissection. Plant J. 44:1054–1064.

[18] 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.

[19] Garris, A.J., S.R.Mc Couch and S. Kresovich. 2003. Population structure and its effects on haplotype diversity and linkage disequilibrium surrounding the xa5 locus of rice OryzasativaL. Genetics 165:759-769.

[20] Glazier A.M., Nadeau J.H. and T.J.Aitman. 2002. Finding genes that underlie complex traits. Science 298:2345-2349. .

[21] Gordon, D., and S.J. Finch. 2005. Factors affecting statistical power in thedetection of genetic association. J. Clin. Invest. 115:1408–1418.

[22] Hackett, C.A. 2002. Statistical methods of QTL mapping in cere-als. Plant MolBiol 48:585-599.

[23] Hamblin, M.T., Mitchell, S.E., White G.M., Gallego J., Kukatla R., Wing R.A., Paterson A.H. and S. Kresovich 2004. Comparative population genetics of the panicoid grasses: Sequence polymorphism, linkage disequilibrium and selection in a di-verse sample ofSorghum bicolor. Genetics 167:471-483.

[24] Hedrick, P.W. 1987.Gametic disequilibrium measures: Proceed with cau-tion. Genetics 117:331–341.

[25] Hunter, K.W., and N.P. Crawford. 2008. The future of mouse QTL map-ping to diagnose disease in mice in the age of whole-genome associa-tion studies. Annu. Rev. Genet. 42:131–141.

[26] Hyten, D.L., Choi I.Y., Song Q., Shoemaker R.C., Nelson R.L., Costa J.M., Specht J.E. and Cregan P.B. 2007. Highly variable pat-terns of linkage disequilibrium in multiple soybean popula-tions. Genetics 175:1937-1944.

[27] Kraakman, A.T.W., Martínez F., Mussiraliev B., van Eeuwijk F.A. and Niks R.E. 2006. Linkage disequilibrium mapping of mor-phological, resistance, and other agronomically relevant traits in modern spring barley cultivars. Mol Breed 17:41-58.

[28] Kraakman, A.T.W., R.E. Niks, Van der Berg P.M.M.M., Stam P. and F.A. Van Eeuwijk. 2004. Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics 168:435-46.

[29] Mackay, T.F. 2001. The genetic architecture of quantitative traits. Annu Rev Genet 35:303-339.

[30] Mackay, I., and W. Powell. 2007. Methods for linkage disequilibrium map-ping in crops. Trends Plant Sci. 12:57–63.

[31] Malysheva-Otto L. and Röder M.S. 2006. Haplotype diversity in the endosperm specificβ-amylase geneβmy1 of cultivated barley (*Hordeumvulgare*L.). Mol Breed 18:143-156.

[32] Oraguzie, N.C., Rikkerink E.H.A., Gardiner S.E. and Silva H.N. de 2007. Association Mapping in Plants. Springer, Tokio and New York, 277 pp.

[33] Palaisa, K.A., M. Morgante, M. Williams, A. Rafalski 2003. Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. Plant Cell,15:1795-1806.

[34] Patterson, N., A.L. Price, and D. Reich. 2006. Population structure and eigenanalysis. PLoS Genet. 2:e190.

[35] Price, A.L., N.J. Patterson, R.M. Plenge, M.E. Weinblatt, N.A. Shadick, et al. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38: 904–909. doi:10.1038/ng1847.

[36] Pritchard, J.K. and N.A. Rosenberg 1999. Use of unlinked genetic markers to detect population stratification in association studies. Am J Hum Genet 65:220-28.

[37] Pritchard, J.K., M. Stephens, N.A. Rosenberg and P. Donnelly 2000. Association mapping in structured populations. Am J Hum Genet 37:170-181.

[38] Pritchard, J., M. Stephens, N.Rosenberg and P. Donnelly. 2000. Association mapping in structured populations. Am. J. Hum. Genet.67:170–181.

[39] Rafalski, J.A. 2010 Association genetics in crop improvement. Curr. Opin. Plant Biol. 13:174–180.

[40] Ravel C., Praud S., Murigneux A., Linossier L., Dardevet M., Balfourier F., Dufour P.H., Brunel D. and G. Charmet.2006. Identification of Glu-B1-1as a candidate gene for the quan-tity of high-molecular-weight glutenin in bread wheat (TriticumaestivumL.) by means of an association study, TheorAppl Genet 112:738-743.

[41] Remington, D.L., Thornsberry, J.M., Matsuoka Y., Wilson L.M., Whitt S.R., Doebley J., Kresovich S., Goodman M.M. and Buckler E.S. IV 2001. Structure of linkage disequilibrium and pheno-typic associations in the maize genome. Proc Nat AcadSci USA 98:11479-11484.

[42] Rosenberg, N.A., Pritchard J.K., Weber J.L., Cann H.M., Kidd K.K., Zhivotovsky L.A. and Feldman M.W. 2002. Genetic structure of human populations. Science 298:2381-2385.

[43] Rostoks, N., Ramsay L., MacKenzie K., Cardle L., Bhat P.R., Roose M.L., Svensson J.T., Stein N., Varshney R.K., D.F. Marshall, et al. 2006. Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop va-rieties. ProcNatlAcadSci USA 103:18656-18661.

[44] Singh, B.D., A.K. Singh. 2016. Marker assisted plant breeding: principles and practices, Springer

[45] Tenaillon, M.I., M.C. Sawkins, A.D. Long, R.L. Gaut, J.F. Doebley, B.S. Gaut 2001. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays*spL.). ProcNatlAcadSci USA,98:9161-9166.

[46] Thornsberry, J.M., Goodman, M.M., Doebley, J., Kresovich, S., Nielsen, D. and Buckler E.S. IV 2001. Dwarf8 polymorphisms associ-ate with variation in flowering time. Nat Genet 28:286-289.

[47] Tommasini, L., Schnurbusch T., Fossati D., Mascher F. and Keller B. 2007. Association mapping of Stagonosporanodorum blotch resistance in modern European winter wheat varieties. TheorAppl Genet 115:697-708.

[48] Visscher, P.M. 2008. Sizing up human height variation. Nat. Genet. 40:489–490.

[49] Xing Y., Frei U., Schejbel B., Asp T. and T. Lübberstedt. 2007. Nucleotide diversity and linkage disequilibrium in 11 expres-sed resistance candidate genes in *Loliumperenne*. BMC Plant Biol 7:43.

[50] Yan, J.B., T. Shah, M.L. Warburton, E.S. Buckler, M.D. McMullen, and J. Crouch. 2009. Genetic characterization and linkage dis-equilibrium estimation of a global maize collection using SNP markers. PLoS One 4:e8451.

[51] Yan, Jianbing Marilyn Warburton, and Jonathan Crouch. 2011. Association Mapping for Enhancing Maize (Zeamays L.) Genetic Improvement. Crop Sci. 51:433–449.

[52] Yu Jianming and Edward S Buckler 2006. Genetic association mapping and genome organization of maizeCurrent Opinion in Biotechnology 2006, 17:155–160.

[53] Zhu, C., and J. Yu. 2009. Nonmetric multidimensional scaling cor-rects for population structure in association mapping with differ-ent sample types. Genetics 182:875–888.

[54] Zhu, C., M. Gore, E.S. Buckler, and J. Yu. 2008. Status and prospects of association mapping in plants. Plant Gen. 1:5–20.

[55] Zhu, C. et al. 2008. Status and prospects of association mapping in plants.Plant Gen.1, 5–20.