Int.J.Curr.Microbiol.App.Sci (2020) *9*(9): 1313-1325

International Journal of Current Microbiology and Applied Sciences
***ISSN: 2319-7706* Volume 9 Number 9 (2020)**
Journal homepage: http://www.ijcmas.com

Review Article https://doi.org/10.20546/ijcmas.2020.909.161

Association Mapping in Crops Plants

Sandeep Kumar Bangarwa*, Ravi Kumawat, Heera lal Barupal, Amit Kumar and Mukesh Kumar Yadav

Department of Genetics and Plant Breeding, Rajasthan College of Agriculture, Maharana Pratap University of Agriculture and Technology, Udaipur-313001 (Raj.), India

**Corresponding author*

Keywords

Association mapping, Crops plants

Article Info

Accepted:
15 August 2020
Available Online:
10 September 2020

A B S T R A C T

Association mapping, a high-resolution method for mapping quantitative trait loci based on linkage disequilibrium, holds great promise for the dissection of complex genetic traits. General understanding of association mapping has increased significantly since its debut in plants. We have seen a more concerted effort in assembling various association-mapping populations and initiating experiments through either candidate-gene or genome-wide approaches in different plant species. In this review, we describe the current status of association mapping in plants, Relation between LD and Association Mapping, QTL Mapping and Association Mapping, Types of Association mapping, Steps in Association Mapping and Benefits and limitations of association mapping.

Introduction

Increasing the efficiency of selection by maximizing the use of desired genetic variation is one of the critical objectives of any breeding programme. Most traits that are important for fitness and agricultural value of plants are quantitative in nature (Yu and Buckler, 2006). Such traits are influenced by many genes, the environment as well as the interactions between genes and the environment. The genetic mapping and molecular characterization of these genes that contribute to the variation of complex traits has the potential to facilitate genome-assisted breeding for crop improvement such as disease resistance, nutrient- use efficiency or biomass production (Holland, 2007).

The principle of mapping a quantitative trait locus (QTL) was first described in the early 20th century by Sax (1923) wherein seed size in bean, a quantitative trait, was associated with seed coat colour, a morphological marker. However, QTL identification did not receive serious attention until the introduction of polymerase chain reaction (PCR)-based molecular markers in the late 1980s. Molecular markers offer great opportunities for dissecting complex traits using QTL mapping. Systematic identification of QTLs in plants was first described by Steven

Tanksley's group while constructing a restriction fragment length polymorphism-based genetic map for tomato fruit quality traits using an interspecific backcross (Paterson *et al.,* 1988). Since then, QTL mapping has been widely used in plants for genetic dissection of biomass, yield, and disease resistance traits. Although there are many published reports on QTLs, only a few QTLs have been used in breeding programmes (Bernardo, 2008), and these are in fact major genes rather than QTLs. Usually, QTL intervals are quite long,; 5–10 cM (wherein 1 cM is equal to 3000 kb of DNA), and contain many genes. Therefore, transferring multiple minor QTLs across genotypes via marker-assisted breeding (MAB) can lead to genetic drag due to presence of undesirable traits in these regions. Often, prior to their utility in positional cloning or MAS in crop improvement efforts, individual genes in QTL regions must be identified via fine-mapping, a time-consuming, laborious, and costly effort. This is due to the necessity of making large numbers of crosses to elicit sufficient numbers of meiotic events. Additionally, QTL identification is based on bi-parental crosses, and quite often QTLs are specific to the bi-parental population used in identifying these QTLs. Thus far, most QTLs identified in either plant or mammalian populations are not useful in a wide range of genetic backgrounds (Sorkheh *et al.,* 2008).

The two most commonly used methods to dissect quantitative traits are linkage mapping and association mapping (Flint-Garcia *et al.,* 2005). Linkage mapping exploits the linked inheritance of functional polymorphisms and adjacent molecular markers within pedigrees of known structures (Mackay, 2001). In plants, such experiments typically are conducted with experimental populations that were derived from crosses of two homozygous genotypes. In contrast, association mapping examines the linked inheritance of functional polymorphisms and adjacent molecular markers in a set of genotypes with unknown ancestry. As the unknown ancestry can extend across thousands of generations, the linked inheritance will only persist for very closely linked polymorphisms. Hence, association mapping exploits historical recombination, whereas linkage mapping makes use of only recombination occurring since the establishment of the mapping population. By exploring deeper population genealogy rather than family pedigrees, association mapping offers mainly three advantages over linkage mapping: higher mapping resolution, higher number of alleles and broader reference population and less research time in establishing an association (Flint-Garcia *et al.,* 2003) (Table 1).

LD in crops

Studies on the rates of decay of LD in various plant taxa (Flint-Garcia *et al.,* 2003), such as maize (Zea mays *subsp*. mays) (Stich *et al.,* 2005; Remington *et al.,* 2001; Tenaillon *et al.,* 2001; Ching *et al.,* 2002; Palaisa *et al.,* 2003 and Gore *et al.,* 2009), barley (Hordeum vulgare) (Caldwell *et al.,* 2004 and Caldwell *et al.,* 2006), Arabidopsis thaliana (Nordborg *et al.,* 2002; Nordbor *et al.,* 2005 and Kim *et al.,* 2007), sorghum (Sorghum bicolor) (Hamblin *et al.,* 2005) and rapeseed (Brassica napus) (Kim *et al.,* 2007), indicate tremendous variation in the extent of LD between various species. This variation is mostly because of the differences in mating type (Nordborg, 2000) and population history. In addition to the variable extent of LD between different crop species, there is also a considerable difference between the germplasm types examined. The population sample effect is evident in maize, where LD decays within 1 kb in land races (Tenaillon *et al.,* 2001), but within approximately 2 kb in

diverse inbred lines (Remington *et al.,* 2001) and can extend up to several hundred kb in commercial elite inbred lines (Ching *et al.,* 2002 and Jung *et al.,* 2004). LD decay can also vary considerably between different genomic regions. For example, significant LD was observed up to 4 kb for the Y1 locus (encoding phytonene synthase), but was seen only over 1 kb for PSY2 (a putative phytonene synthase) in the same maize population (Palaisa *et al.,* 2003).

Relation between LD and Association Mapping

LD refers to the non-random association of alleles at different loci. Both linkage and association mapping approaches are based on the LD between molecular markers and functional loci. However, the difference between both approaches is that in linkage mapping the LD used is generated by the mating design, while in association mapping the LD used is present in the germplasm set under study. Furthermore, given the balanced design, LD is influenced only by recombination in linkage mapping populations in the absence of segregation distortion. Therefore, the extent of LD in such a population does not have to be inferred empirically, but can be derived theoretically based on information regarding the population type. However, in an association mapping context, LD might not only be influenced by recombination but also by various other forces (Flint-Garcia *et al.,* 2003). These forces influencing the pattern and extent of LD are: (i) mating type, (ii) genetic drift, (iii) selection, (iv) mutation, (v) population substructure and relatedness and (vi) ascertainment bias (Clark *et al.,* 2005; Stich *et al.,* 2005 and Yu *et al.,* 2006). Given that for most association mapping populations, the importance of these factors is unknown or can only be roughly estimated, the extent of LD cannot be derived theoretically but must be inferred from empirical studies based on molecular markers in order to evaluate the applicability and resolution of association mapping approaches. However, in order to interpret the results of empirical studies of LD, knowledge about the properties of the applied measures of LD is of crucial importance.

QTL Mapping and Association Mapping

Association mapping (AM), also known as linkage disequilibrium (LD) mapping, has been proposed as an alternative approach to overcome limitations of pedigree based QTL mapping. In AM, genotype and phenotype correlations are investigated in unrelated individuals. Unlike QTL mapping, AM takes advantage of LD as well as historical recombinations present within the gene pool of an organism, thus utilizing a broader reference population (Breseghello and Sorrels, 2006a; Ersoz and Buckler, 2007; Myles *et al.,* 2009). If two alleles from separate loci occur together more often than otherwise predicted, on the basis of their individual frequencies, i.e. non-random association of alleles at separate loci, they are deemed to be in LD. Only those molecular markers that are tightly linked to the trait and located within the extent of LD decay will demonstrate significant marker-trait association. If markers are not tightly linked to a trait, they will be separated by recombination during meiosis throughout the evolutionary history of the crop. Accumulating meiotic events in a population will increase the statistical power and mapping resolution for detecting associations. However, it should be noted that the rate of LD decay should be sufficient enough to statistically identify associations, but not too high as it will make it difficult to narrow down the target genomic region. AM requires availability of large numbers of polymorphic markers and is more complex than QTL

mapping, as historical factors such as population admixture, selection, and genetic drift can bias the detected association. Moreover, the population genetic structure as well as effects due to non-random mating (relatedness) must be accounted for in the analysis to avoid false positive (spurious) associations. Population structure influences both the power and precision of detecting associations. However, it can be overcome with good sampling and by using appropriate algorithms to detect groupings in a population and accounting for these in an association mapping analysis (Zhu *et al.,* 2008). Early on, LD mapping had been used in human studies to understand the genetic control of disease. Nowadays, it has rapidly gained interest among plant scientists for studying biomass traits, yield, and disease resistance, among others. Reviews on the concept, methodology, prospects, and status of LD and AM studies in plants have already been published (Neale and Savolainen, 2004; Gupta *et al.,* 2005; Breseghello and Sorrels, 2006a; Ersoz and Buckler, 2007; Abdurakhmonov and Abdukarimov, 2008; Sorkheh *et al.,* 2008; Zhu *et al.,* 2008; Myles *et al.,* 2009; Neale and Kremer, 2011).

Types of Association mapping

Based on the scale and focus of a particular study, association mapping generally falls into two broad categories,

Genome wide association mapping and
Candidate gene association mapping.

Genome wide association mapping

Genome-wide association mapping or genome scan, which surveys genetic variation in the whole genome to find signals of association for various complex traits (Risch and Merikangas, 1996). Basically, entire genome is scanned using SNP or SSR rich marker, with saturated genome map leads to maximum opportunities for marker trait association and finding all possible across entire genome under study.

Currently, various researches are going on with various plants for different agronomic, quality, yield, abiotic and biotic stress resistant traits. Recently, Laido *et al.,* 2014 has studied whole genome association mapping analysis in wheat by employing 230 inbred line and phenotype various trait like plant height (PH), heading date (HD), protein content (PC) and thousand kernel weight (TKW). This study shows, in most of case of wheat population, r2 fall in range of 0.12 to 0.18. This implies that with few SNP marker its possible to scan entire genome. Study identified 89 QTLs for these traits, from 25 QTLs for PH, to 42 QTLs for TKW 44 QTLs concur with previously mapped QTLs with genes for photoperiod insensitivity (ppd), vernalisation (vrn), earliness per se(eps), dwarfism (rht) and grain protein content (gpc).Belo *et al.,* 2008 scan whole genome and detect alleleic variant for *fad2* associated with increased oleic acid level with maize using GWAS. Study detected unique MTA at MZA10924 loci on chromosome 4, in 4 out of 7 sub population. Loci are found to be linked with *fad2* gene.

Using of promising GWAS strategies, allows development of saturated linkage maps and highly informative microsatellite and SNP markers in plants makes it possible to systematically survey marker-trait association on a whole-genome scale gives opportunity for fine mapping. (Soto-Cerda, 2012) various outcome and potent utilization of GWAS method for fine gene mapping has been shown in table 1. As new variant MAGIC (Multiparent Advanced Generation Intercross) method has provide has the ability to capture the majority of the variation available in the gene pool. Although it might take several

years for constructing population, these populations are suitable for fine mapping, (Mott *et al.,* 2000).

Candidate gene association mapping

CGs refers to either cloned genes presumed to affect a given trait ('functional CGs') or to genes suggested by their close proximity on linkage maps to loci controlling the trait ('positional CGs'). CG analysis is based on the hypothesis that known-function genes (the candidate genes) could correspond to loci controlling traits of interest steps: Identification of CG, calculate statistical correlations between CG polymorphisms and phenotypic variation in a set of genealogically unrelated individuals, checking marker trait co-segregation and fine mapping of QTL and finally validation of CG. Recently, Syngenta have introduced drought tolerant Artesian hybrid maize to combat effect of drought. Previously, Setter *et al*, 2010 using CG approach studied drought tolerance for maize in watered plot and stressed case-control study detected putative MTA of SNP 186 with aldehyde oxidase gene *ZmAO3*.

Steps in Association Mapping

Association Mapping Population

The choice of germplasm is crucial to the success of association analysis (Flint-Garcia *et al.,* 2005 and Breseghello *et al.,* 2006). This is because genetic diversity, the extent of genome-wide LD, as well as the level of population structure and relatedness in the population under consideration determine the mapping resolution, the appropriate method for association analysis and the statistical power to detect marker-phenotype associations. The following possibilities have been proposed to classify association mapping populations. Firstly, plant populations can be classified with respect to the two dimensions: (a) the extent of population structure and (b) familial relatedness (Yu and Buckler, 2006). These authors proposed that with respect to these two dimensions, all populations are classifiable into one of the five groups: (i) ideal sample with subtle population structure and familial relatedness, (ii) multi-family sample without population structure, (iii) sample with population structure but without familial relatedness, (iv) sample with both population structure and familial relatedness and (v) sample with severe population structure and familial relatedness. Due to local adaptation, selection, and breeding history in many plant species, many populations for association mapping would fall into category four. Alternatively, we can classify populations according to the source of materials, germplasm bank collections, synthetic populations, and elite germplasm (Breseghello and Sorrells, 2006a). The assignment of a population to one of these classes determines which statistical methods have to be applied later on for association analysis. Alternatively, association mapping populations can be classified with respect to the source of materials (Breseghello and Sorrells, 2006). Four main types of populations can be considered in a plant genetics context: (i) natural populations, (ii) germplasm bank collections, (iii) synthetic populations and (iv) elite breeding material. These four types of populations are expected to differ considerably with respect to the following aspects: (i) genotypic and phenotypic diversity, (ii) extent of LD and (iii) importance of population structure and familial relatedness.

Genotyping for association mapping

In association mapping studies, genotyping is required for inferences both on the population structure and relatedness as well as on marker-phenotype associations. With respect to the first task, the genotyping of a set of

selectively neutral background markers distributed throughout the genome is required. Random amplified polymorphic DNA (RAPD) and amplified fragment length polymorphism (AFLP) markers can serve as background markers, but as a result of their dominant inheritance demand special statistical methods if used to estimate population genetic parameters (Falush *et al.,* 2007). Conversely, codominant simple sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs) are more powerful in estimating population structure and familial relatedness (Ritland, 2005). Because SSR markers are multiallelic, reproducible and mostly selectively neutral (Smith *et al.,* 1997 and Vignal *et al.,* 2002), they have been the predominant molecular marker to study kinship and population structure. Given their higher genome density, lower mutation rate and better amenability to high throughput detection systems, SNPs are rapidly becoming the markers of choice for complex trait dissection studies (Hamblin *et al.,* 2007). On a per-site basis, SNPs are, because of their predominantly biallelic nature, less informative than multiallelic SSRs.

Phenotyping for association mapping

Efficient field design with incomplete block design (e.g.,α-lattice), appropriate statistical methods (e.g., nearest neighbor analysis and spatial models), and consideration of QTL × environmental interaction should be explored to increase the mapping power, particularly if the field conditions are not homogenous (Eskridge, 2003). Association mapping studies often are long-term projects, with phenotyping being conducted over years in multiple locations (Flint-Garcia *et al.,* 2005). In this framework, any newly discovered candidate gene polymorphism can always be tested for association with existing phenotypic data. Also, transitioning from a candidate-gene to a genome-wide approach should be seamless if the original association mapping panel was constructed in a manner such that other complex traits can be evaluated and robust phenotypic data were collected along the way. To ensure that high quality data are obtained from a wide range of conducted experiments, each researcher should assess the quality of the experiment for which they are responsible. Specific information about the experiment, such as check performance and environmental growth conditions (field or greenhouse), should be included as an annotation to the experiment in the trait database. In the case that unbalanced plant breeding trials are used as sources of phenotypic data, the proper statistical modelling of the experimental design and especially the consideration of genotype-environment as well as marker-environment interactions (Malosetti *et al.,* 2008) increases the mapping power (Stich *et al.,* 2008).

Statistical analysis

The basic statistics for association analysis, under an ideal situation, would be linear regression, analysis of variance (ANOVA), *t* test or chi-square test. For family-based samples, the transmission disequilibrium test (TDT) (Spielman *et al.,* 1993) is used to study the genetic basis for human disease, whereas the quantitative transmission disequilibrium test (QTDT) is employed in the dissection of quantitative traits (Abecasis *et al.,* 2000; Allison, 1997). To address the issue of population structure in population-based samples, GC and SA are the two most common methods utilized in both human and plant association studies. With GC, a set of random markers is used to estimate the degree that test statistics are inflated by population structure, assuming such structure has a similar effect on all loci (Devlin and Roeder, 1999). By contrast, SA analysis first uses a set of random markers to estimate population structure (Q) and then incorporates this

estimate into further statistical analysis (Falush *et al.*, 2003; Pritchard and Rosenberg, 1999; Pritchard *et al.*, 2000a). A unified mixed-model approach for association mapping that accounts for multiple levels of relatedness was recently developed (Yu *et al.*, 2006). In this method, random markers are used to estimate Q and a relative kinship matrix (K), which are then fit into a mixed model framework to test for marker-trait associations. As this mixed-model approach crosses the boundary between family-based and population-based samples, it provides a powerful complement to currently available methods for association mapping (Zhao *et al.*, 2007). Principal component analysis (PCA) has long been used in genetic diversity analysis and was recently proposed as a fast and effective way to diagnose population structure (Patterson *et al.*, 2007; Price *et al.*, 2006). TASSEL is the most commonly used soft ware for association mapping in plants and is frequently updated as new methods are developed (Bradbury *et al.*, 2007).

In addition to association analysis methods (i.e., logistic regression, linear model, and mixed model), TASSEL is also used for calculation and graphical display of linkage disequilibrium statistics and browsing and importation of genotypic and phenotypic data. STRUCTURE soft ware typically is used to estimate Q (Pritchard *et al.*, 2000a). The Q is an $n \times p$ matrix, where n is the number of individuals and p is the number of defined subpopulations. SPAGeDi soft ware is used to estimate K among individuals (Hardy and Vekemans, 2002). K is an $n \times n$ matrix with off –diagonal elements being *Fij*, a marker-based estimate of probability of identity by descent.

The diagonal elements of K are one for inbreds and $0.5 \times (1 + Fx)$ for non-inbred individuals, where Fx is the inbreeding coefficient. EINGENSTRAT software is used to estimate PCs of the marker data and correct test statistics resulting from population stratification (Price *et al.*, 2006). Other software commonly used in human association mapping includes Merlin (Abecasis *et al.*, 2002) and QTDT (Abecasis *et al.*, 2000). SAS software (SAS Institute, 1999) or R (Ihaka and Gentleman, 1996) often are used by advanced researchers with programming skills as the platform to develop various methods. ASREML (Gilmour *et al.*, 2002) and MTDFREML (Boldman *et al.*, 1993) are two of several software packages used in animal genetics in mixed model analysis of data from a very large number of individuals (Table 2).

Table.1 Comparison of the properties of different methods to dissect quantitative traits

Properties	Linkage Mapping	Nested Association Mapping	Association Mapping
Resolution of mapping	Low	Medium	High
Time to establish the required germplasm	Medium	Medium	Low
Number of alleles evaluated	2	>20	>40
QTL detection in genome-scan	High	High	Low
Relevance of population substructure	No	No	High
Genotyping requirement for genome-scan	Low	Medium	High

Table.2

S.No	Organism / Species	Character of study	References
1	Arabidopsis	Flowering time, Pathogen resistance genes	Aranzana *et al.*, 2005
		Downy mildew resistance genes	Li *et al.*,2010
		Climate sensitive QTL	Nemri *et al.*, 2010
2	Maize	Northern leaf blight (NLB) resistance	Kump *et al.*, 2011
		Southern leaf blight (SLB) resistance	Poland *et al.*, 2011
		Leaf architecture	Tian *et al.*, 2011
3	Barley	Putative anthocyanin pathway gene, *HvbHLH1*	Cockram *et al.*, 2010
4	Wheat	Abiotic stress aluminium resistance	Raman *et al.*, 2010
		Flowering time	Rousset *et al.*, 2011
5	*Brassica napus*	Seed oil related loci	Zou *et al.*, 2010

Benefits and limitations of Association mapping

The potential high resolution in localizing a QTL controlling a trait of interest is the primary advantage of AM as compared to linkage mapping. 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 and 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 and 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 under powering 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 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 and 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 and Zeggini, 2010), several new association analysis methods for rare variants are being proposed, with some important advances in complex trait dissection (Li and Leal, 2008).

References

Abecasis, G.R., S.S. Cherny, W.O. Cookson, and L.R. Cardon. 2002. Merlin–rapid analysis of dense genetic maps using sparse gene flow trees. Nat. Genet. 30:97–101.

Abdurakhmonov IY, Abdukarimov A. 2008. Application of association mapping to understanding the genetic diversity of plant germplasm resources. International Journal of Plant Genomics 2008,574–927.

Allison, D.B. 1997. Transmission-disequilibrium tests for quantitative traits. Am. J. Hum. Genet. 60:676–690.

Aldwell KS, Russell J, Langridge P, Powell W. Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, Hordeum vulgare. Genetics 2006;172:557–67.

Aranzana MJ, Kim S, Zhao KY, *et al.*, 2005. Genome-wide association mapping in Arabidopsis identifies previously known flowering time and pathogen resistance genes. PLoS Genetics 1,531–539.

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

Belo, A., P. Zheng, S. Luck, B. Shen, D.J. Meyer, B. Li, S. Tingey, and A. Rafalski. 2008. Whole genome scan detects an allelic variant of fad2 associated with increased oleic acid levels in maize. Mol. Genet. Genomics 279:1–10.

Bernardo R. 2008. Molecular markers and selection for complex traits in plants: learning from the last 20 years. Crop Science 48, 1649–1664.

Brachi, B., Faure, N., Horton, M., Flahauw, E., Vazquez, A., Nordborg, M., Bergelson, J.,Cuguen, J. and 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. and Buckler, E. (2007).TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, Vol. 23, No. 19, 2633-2635.

Breseghello, F., and M.E. Sorrells. 2006a. Association analysis as a strategy for improvement of quantitative traits in plants. Crop Sci.46:1323–1330.

Boldman, K.G., L.A. Kriese, L.D. Van Vleck, and S.D. Kachman. 1993. A manual for the use of MTDFREML: A set of programs to obtain estimates of variances and covariances. ARS, USDA, Washington, DC.

Caldwell KS, Langridge P, Powell W.

Comparative sequence analysis of the region harboring the hardness locus in barley and its colinear region in rice. Plant Physiology 2004; 136: 3177–90.

Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, *et al.*, SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genetics 2002; 3:19.

Clark AG, Hubisz MJ, Bustamante CD, Williamson SH, Nielsen R. Ascertainment bias in studies of human genome wide polymorphism. Genome Research 2005; 15:1496–502.

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. and 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.

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

Ersoz ES, Yu J, Buckler ES. 2007. Applications of linkage disequilibrium and association mapping in crop plants. Springer, pp 97–119.

Eskridge, K.M. 2003. Field design and the search for quantitative trait lociin plants. Available at: http://www.stat.colostate.edu/graybillconference2003/Abstracts/Eskridge.html; verifi ed 20 May 2008.

Flint-Garcia, S., Thornsberry, J. and Bukler, E. (2005). Structure of linkage disequilibrium in plants. *Annual Review Plant Biology*, Vol. 54, 357-374.

Flint-Garcia SA, Thuillet AC, Yu J, Pressoir G, Romero SM and Mitchell SE. 2005. Maize association population: a high resolution platform for quantitative trait locus dissection. Plant Journal, 44:1054–64.

Falush D, Stephens M. and Pritchard J.K. 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. Molecular Ecology Notes, 7:574–8.

Falush, D., M. Stephens, and J.K. Pritchard. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics 164:1567–1587.

Gilmour, A.R., B.J. Gogel, B.R. Cullis, S.J. Welham, and R. Th ompson.2002. ASReml user guide release 1.0. VSN International Ltd., Hemel Hempstead, UK.

Gore MA, Chia JM, Elshire RJ, Sun Q, Ersoz ES, Hurwitz BL, *et al.*, A first-generation haplotype map of maize. Science 2009; 326: 1115–7.

Gupta PK, Rustgi S, Kulwal PL. 2005. Linkage disequilibrium and association studies in higher plants: present status and future prospects. *Plant Molecular Biology* 57, 461–485.

Hamblin MT, Salas Fernandez MG, Casa AM, Mitchell SE, Paterson AH, Kresovich S. Equilibrium processes cannot explain high levels of short- and medium-range linkage disequilibrium in the domesticated grass Sorghum bicolor. Genetics 2005; 171: 1247–56.

Hamblin MT, Warburton ML, Buckler ES. Empirical comparison of Simple Sequence Repeats and single nucleotide polymorphisms in assessment of maize diversity and relatedness. PLoS One 2007;2(12):e1367.

Hardy, O.J., and X. Vekemans. 2002. SPAGeDi: A versatile computer programto analyse spatial genetic structure at the individual or population

levels. Mol. Ecol. Notes 2:618–620.

Holland JB. Genetic architecture of complex traits in plants. Current Opinion in Plant Biology 2007; 10: 156–61.

Ihaka, R., and R. Gentleman. 1996. R: A language for data analysis and graphics. J. Comput. Graph. Stat. 5:299–314.

Jung M, Ching A, Bhattramakki D, Dolan M, Tingey S, Morgante M, *et al.*, Linkage disequilibrium and sequence diversity in a 500-kbp region around the adh1 locus in elite maize germplasm. Theoretical and Applied Genetics 2004; 109: 681–9.

Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S, *et al.*, Recombination and linkage disequilibrium in Arabidopsis thaliana. Nature Genetics 2007; 39: 1151–5.

Kump, K., Bradbury, P., Wisser, R., Buckler, E., Belcher, A., Oropeza-Rosas, M., Zwonitzer, J., Kresovich, S., McMullen, M., Ware, D., Balint-Kurti, P. and 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.

Li, Y., Huang, Y., Bergelson, J., Nordborg, M. and Borevitz, J. (2010). Association mapping of localclimate-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.

Li, B. and 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.

Mackay TF. The genetic architecture of quantitative traits. Annual Review of Genetics 2001; 35:303–39.

Mott, R., and J. Flint. 2002. Simultaneous detection and fi ne mapping of quantitative trait loci in mice using heterogeneous stocks. Genetics 160:1609–1618.

Malosetti M, Ribaut JM, Vargas M, Crossa J, van Eeuwijk FA. A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress trials in maize (*Zea mays* L.). Euphytica 2008; 161: 241–57.

Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES. 2009. Association mapping: critical considerations shift from genotyping to experimental design. *The Plant Cell* 21, 2194–2202.

Neale DB, Savolainen O. 2004. Association genetics of complex traits in conifers. Trends in Plant Science 9, 325–330.

Nemri, A., Atwell, S., Tarone, A., Huang, Y., Zhao, K., Studholme, D., Nordborg, M. and 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.

Neale DB, Kremer A. 2011. Forest tree genomics: growing resources and applications. Nature Reviews Genetics 12, 111–122.

Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J, Hagenblad J, *et al.*, The extent of linkage disequilibrium in Arabidopsis thaliana. Nature Genetics 2002;30:190–3.

Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng H, *et al.*, The pattern of polymorphism in Arabidopsis thaliana. PLoS Biology 2005; 3(7): e196. doi:10.1371/ journal.pbio.0030196.

Nordborg M. Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. Genetics 2000; 154: 923–9.

Poland, J., Bradbury, P., Buckler, E. and 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.

Paterson, A. H., Lander E. S., Hewitt J D, Peterson S, Lincoln S E, Tanksley S D. 1988.Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335, 721–726.

Patterson, N., A.L. Price, and D. Reich. 2007. Population structure andeigenanalysis. PLoS Genet 2:e90.

Price, A.L., N.J. Patterson, R.M. Plenge, M.E. Weinblatt, N.A. Shadick and D. Reich. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38: 904–909.

Palaisa, K., Morgante, M., Tingey, S. and 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.

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

Pritchard, J.K., and N.A. Rosenberg. 1999. Use of unlinked genetic markersto detect population stratifi cation in association studies. Am. J.Hum. Genet. 65:220–228.

Raman, H., Stodart, B., Ryan, P., Delhaize, E., Emebiri, L., Raman, R., Coombes, N. and 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.

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, *et al.*, Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proceedings of the National Academy of Sciences, USA 2001;98:11479–84.

Risch, N., and K. Merikangas. 1996. Th e future of genetic studies of complex human diseases. Science 273:1516–1517.

Ritland K. Multilocus estimation of pairwise relatedness with dominant markers. Molecular Ecology 2005;14:3157–65.

Rousset, M., Bonnin, I., Remoué, C., Falque, M., Rhoné, B., Veyrieras, J., Madur, D.,Murigneux, A., Balfourier, F., Le Gouis, J., Santoni, S. and 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.

Sax K. 1923. The association of size differences with seed-coatpattern and pigmentation in *Phaseolus vulgaris*. Genetics 8, 552–560.

Smith JSC, Chin ECL, Shu H, Smith OS, Wall SJ, Senior ML, *et al.*, An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.): comparisons with data from RFLPS and pedigree. Theoretical and Applied Genetics 1997; 95: 163–73.

Sorkheh K, Malysheva-Otto LV, Wirthensohn MG, Tarkesh-Esfahani S¸ Martı´nez-Go´ mez P. 2008. Linkage disequilibrium, genetic association mapping and gene localization in crop plants. Genetics and Molecular Biology 31, 805–814.

Spielman, R., McGinnis, R. and 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.

Stich, B. and 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.

Stich B, Melchinger AE, Heckenberger M, Mohring J, Schechert A, Piepho HP. Association mapping in multiple segregating populations of sugar beet (*Beta vulgaris* L.). Theoretical and Applied Genetics 2008; 117: 1167–79.

Stich B, Melchinger AE, Frisch M, Maurer HP, Heckenberger M, Reif JC. Linkage disequilibrium in European elite maize germplasm investigated with SSRs. Theoretical and Applied Genetics 2005; 111: 723–30.

Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. mays L.). Proceedings of the National Academy of Sciences, USA 2001; 98: 9161–6.

Tian, F., Bradbury, P., Brown, P., Hung, H., Sun, Q., Flint-Garcia, S., Rocheford, T., McMullen, M., Holland, J. and 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.

Vignal A, Milan D, SanCristobal M, Eggen A. A review on SNP and other types of molecular markers and their use in animal genetics. Genetics, Selection, Evolution 2002; 34: 275–305.

Yu, J., M. Arbelbide, and R. Bernardo. 2005. Power of *in silico* QTL mapping from phenotypic, pedigree, and marker data in a hybrid breeding program. Theor. Appl. Genet. 110:1061–1067.

Yu, J. and 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 WH, VrohBi I, Yamasaki M, Doebley JF, *et al.*, A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics 2006; 38: 203–8.

Zhang, D., Bai, G., Zhu, C., Yu, J. and 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., M.J. Aranzana, S. Kim, C. Lister, C. Shindo, C. Tang, C. Toomajian,H. Zheng, C. Dean, P. Marjoram, and M. Nordborg. 2007. Anarabidopsis example of association mapping in structured samples. PLoS Genet 3:e4.

Zhu, C., Gore, M., Buckler, E. and 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. and Meng, J. (2010). Association mapping ofseed oil content in *Brassica napus* and comparison with quantitative trait loci identified from linkage mapping. *Genome*, Vol. 53, No. 11, 908-916.

How to cite this article:

Sandeep Kumar Bangarwa, Ravi Kumawat, Heera lal Barupal, Amit Kumar and Mukesh Kumar Yadav. 2020. Association Mapping in Crops Plants. *Int.J.Curr.Microbiol.App.Sci.* 9(09): 1313-1325. doi: https://doi.org/10.20546/ijcmas.2020.909.161