Linkage disequilibrium mapping: A journey from traditional breeding to molecular breeding in crop plants

Jyoti Prakash Sahoo[a,*], Upasana Mohapatra[b], Debanjana Saha[c], Iswar Chandra Mohanty[a] & Kailash Chandra Samal[a]

[a]Department of Agricultural Biotechnology, OUAT, Bhubaneswar 751 003, Odisha, India

[b]Department of Plant Biotechnology, UAS, GKVK, Bengaluru 560 065, Karnataka, India

[c]Department of Biotechnology, CUTM, Bhubaneswar 752 050, Odisha, India

E-mail: jyotiprakashsahoo2010@gmail.com

Germplasms are the reservoir of agronomically important traits traditionally maintained by various tribal communities over the year. Maintaining these germplasms generations after generations has little value unless exploited for the desired agronomic traits like biotic and abiotic stress, yield attributes and nutritional enrichment. Association mapping, otherwise called linkage disequilibrium mapping, is a molecular breeding approach for characterizing complex traits with agronomic importance in crop plants. It is a systematic method for identifying novel traits and is treated as an alternative tool to traditional QTL mapping approaches, which involves correlating molecular markers with the phenotypic trait in a diversified population. The map's resolution in association mapping is based on the candidate-gene approach or genome-wide association approach. Therefore, association mapping studies offer a great perspective on crop genetic improvement. Still, considerably large-scale research is required to determine the sensible implementation of association mapping analysis in most crop plants. Currently, there is considerable interest in using association mapping approaches in crop breeding programs, which can be achieved by advanced genomic technology and the development of statistical computer software packages. Here, the linkage disequilibrium approach and its usefulness in association mapping studies, including the steps associated with it are discussed. The current status and future challenges in complex trait dissection by utilizing the linkage disequilibrium mapping in crop plants are also discussed.

Keywords: Association mapping, GWAS, Linkage disequilibrium, Molecular markers, QTLs

IPC Code: Int Cl[22]: A01H 1/00, A01H 1/04

Agriculture is the backbone of human civilization. For global food security, exploitation of the crop genetic diversity is a vital requirement to minimize the feeding problem. Worldwide the germplasm resources are generally composed of wild cultivars with natural genetic variability as they are exposed to environmental stresses and selected through crop domestication[1]. Germplasm resources having desired agronomically relevant characteristics have traditionally been kept by various tribal communities, year after year, generation after generation. Therefore, it is of less use until appropriately utilized and explored scientifically. So, it is essential to identify the characters for agronomically important traits through traditional breeding or molecular breeding protocols like linkage mapping or linkage disequilibrium mapping for improvement of the germplasms. The agronomically significant variations in crop plants, such as resistance to biotic and abiotic stresses and high productivity, are controlled by polygenic inheritance of complex traits generally referred to as QTLs (Quantitative Trait Loci), which depend on genetic and environmental factors interaction[2]. Mapping and characterizing these genomic regions in the plant genome can facilitate marker-assisted breeding for crop improvement towards tolerance to various abiotic and biotic stresses[3]. Traditional linkage mapping distinguishes impartially acquired markers near the genes that control the complex quantitative attributes.

This whole approach is made in two different ways, i.e., using a bi-parental mapping population or using lines from a population with a natural genetic variation or by collecting germplasm. The former is designated as QTL mapping[4] and the later is called linkage disequilibrium mapping or association mapping. Linkage mapping dissects the inheritance of

linked polymorphism with the adjacent molecular marker locus where the experimental populations are the crossing lines of two homozygous genotypes with known ancestry[5]. In the case of the association mapping approach, the marker and trait association is exploited in a set of genotypes with unknown ancestry[5]. The details of the conventional QTL-mapping approach have been looked after recently and for the itemized idea, models and techniques, issues and points of view of linkage investigation are proposed[6-8].

Association mapping, otherwise called linkage disequilibrium mapping, aims to increase the map density concerning alleles in a less time duration of research by exploring the family relatedness in a genetically diversified population rather than exploiting the family pedigree[9]. Compared to the conventional linkage mapping, linkage disequilibrium (LD) mapping, using the nonrandom associations of loci in haplotypes, is a powerful high-resolution mapping tool for complex quantitative traits. The recent advances in developing an unbiased association mapping approach for plant population with their successful applications in dissecting several simples to complex traits in many crop species demonstrate a flourish of the approach as a "powerful gene tagging" tool for crops in the plant genomics era of 21st century. Association mapping may significantly increase the efficiency and effectiveness of breeding compared to conventional breeding. The fundamental advantages of association mapping compared to traditional phenotypic selection are simpler than phenotypic screening. This review discusses the linkage disequilibrium approach and its relation with association mapping, steps associated with association analysis, recent studies in crop plants related to marker-trait association, and possible limitations of this approach, including their future perspectives.

Concept of association mapping

Association mapping is a mapping system of the QTLs in plants based on the concept of linkage disequilibrium having an output of a highly resolute map. It is a very prominent and useful tool to dissect the allelic variations directly related to the complex agronomic traits. The association analysis is entirely dependent on the size of the population used in the marker-trait association experiment, the molecular marker density taken, the decay of linkage disequilibrium between the marker and the target trait

and the statistical error in the phenotyping and genotyping analyzed data[10,11]. A scheme chart of the general protocol involved in the association mapping analysis is represented in Figure 1, which is generally followed to identify the novel genes or candidate genes in plants. Association mapping is found as an efficient plant breeding tool that is used widely by plant breeders[12,13]. It is a tool to detect the rare allelic variation in the population. It can be used as an alternative molecular mapping tool to the traditional QTL mapping approach[14]. In a broad sense, association mapping studies are divided into two categories, i.e., candidate gene approach and GWAS (Genome-Wide Association) mapping.

The candidate gene approach is a hypothesis-based approach used to dissect the alleles linked to the complex agronomically important traits. It involves the phenotyping followed by the genotyping methodology. The genotypes present in a germplasm panel are considered to sequence the genes entirely or partially to find out the significant association that causes polymorphism. It is an expensive methodology, but it may help to find rare polymorphism. In the case of genome-wide association analysis, the whole genome of the plant is explored to find out the genetic variation linked to the complex agronomic traits[15]. With the recent development of high-throughput genotyping and phenotyping technologies, genetic variation in many model plants such as rice and maize is being discovered on a genome-wide scale[16], which has great potential to identify risk factors for complex traits in plants. However, only a limited number of recombination events are observed in the QTL mapping approach, whereas with association mapping, the historical recombination events can be
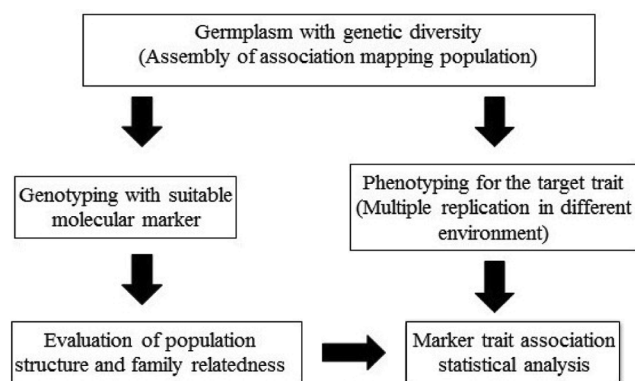

Fig. 1 — The flowchart of the steps associated with association mapping

exploited[17,18]. A comparative study is presented in Table 1 to understand better the candidate gene approach, genome-wide association mapping, and QTL mapping. Nested association mapping is another powerful statistical analysis tool of the marker-trait association approach where recombinant inbred lines are generally used with a small number of single nucleotide polymorphism markers for genotyping[18].

Theory of Linkage disequilibrium and LD decay

The association between the molecular marker loci with the target phenotypic trait is referred to as association mapping. On the other hand, linkage disequilibrium is the association between the two marker alleles in non-random manner[19]. Linkage, which is generally confused with the concept of linkage disequilibrium, implies the inheritance of loci correlated with a known physical distance in a chromosome. However, the correlation analysis within the alleles present in a population is referred to as linkage disequilibrium[20]. In the traditional linkage mapping approach, LD is calculated using the bi-parental mating mapping population. In the case of association mapping, the collected germplasms are used as the mapping population to calculate the linkage disequilibrium[21]. The concept of LD[22] was first explained in 1917 by Jennings, and the procedure of LD quantification was demonstrated in 1964 by Lewontin.

Linkage disequilibrium (D) is the nonrandom association of alleles in a population at different loci in the genome. The mathematical derivation[17] of D is $D = P_{AB} - P_A P_B$. In this equation, P_{AB} is the gametic allele A and B frequency at two loci. P_A and P_B are the frequency products of the allele A and B. There are other two important statistics available to explain D, i.e., D' and r^2, where D' reflects the recombination history and r^2 summarize the history of the recombination and mutation[20,23]. The r^2 value of 0.1 is considered the minimum significant threshold frequency between the observed pair of loci[17]. LD decay is the decline in the magnitude of LD between two loci due to recombination between them. The rate of LD decay is calculated by the product of r and d, where 'r' refers to the frequency of recombination between the two loci and 'd' refers to disequilibrium[21-23]. LD between two loci decays both temporally as the generation advances and spatially with increasing distance between two loci.

Factors affecting the variation in linkage disequilibrium analysis

The linkage disequilibrium analysis performed to find the marker-trait association is generally affected by recombination, the mating system in plants, germplasm and the evolutionary factors such as population structure and genetic drift.

Effect of recombination

LD is strongly affected and influenced by the recombination rate during the localization of the alleles. It is generally associated with the GC content in the genome[24]. GC-rich sequences present in the genome may be associated with high recombination or mutation rates which directly lower the significant levels of LD[25]. In plant genomes like barley, maize,

Table 1 — A comparative study between the QTL mapping and Association mapping approaches

	QTL Mapping	Candidate Gene Approach	GWAS
Essential requirements	Less population size	Large population size	Large population size
	Mapping population (Early mapping populations, RIL, NIL, DH, BIL, etc.) should be generated by the biparental crossing protocol	Less markers density across the genome	Large number of markers
		The genetic and biochemical information regarding the dissecting trait	The genetic and biochemical information regarding the dissecting trait
	Less number of markers	LD and population structure analysis	Linkage disequilibrium and Population structure
Important advantage	No effect of population structure and family relatedness	Fine mapping	Fine mapping
	Rare allele identification	Relatively low cost	Detection of allelic differences
Important disadvantage	Limitation of genetic diversity	Functional knowledge on the trait of interest	May miss to detect the rare alleles
	Unable to find out the difference between pleotropic genes and physically related genes		Population structure analysis

(GWAS – Genome-wide association study, RIL – Recombinant Inbred Lines, NIL – Nearly Isogenic Lines, DH – Double Haploids, BIL – Backcross Inbred Lines, LD – Linkage disequilibrium)

and tomato, inter-chromosomal LD variation has revealed such evidence[26].

The type of mating system

The selfing process reduces the recombination events by increasing the homozygosity in the individuals of the population; therefore, the mating system has a huge impact on linkage disequilibrium[20]. Studies reviled that LD is more in the self-pollinated species like the model plant Arabidopsis, rice and wheat than the cross-pollinated species like rye[27] and maize[28]. These experimental findings suggest an expectation of a fine and high-resolution map in the path of potential candidate gene cloning.

Type of germplasm

The type of germplasm used as the mapping population plays a vital role in the variation of LD as it exploits the genetic diversity. The more in the genetic variation reveals the quick decay in LD. For example, LD decay was observed in maize within 1 kilo base map distance in landraces and approximately 2 kilo base map distance in diversified inbreeds[27,28]. There has been no large-scale assessment of LD in *O. sativa*, although the first study in rice found an LD decay of ~100 kb around a disease resistance locus in the aus subpopulation and a more recent study reported an LD decay of ~50 kb in indica rice and of ~5 kb in *O. rufipogon*[25]. So, the germplasm type has a significant impact on association mapping analysis[25].

Structure of the mapping population

Germplasm selection affects the variation of LD in the genome as locus-specific manner, whereas population structure affects LD throughout the whole genome by exploiting historical recombination events[29]. A significant but spurious LD was reported in maize Dwarf 8 gene and variation in flowering time trait[30] due to the population structure and family relatedness.

Genetic drift

The process of gene flow can create LD between intently related loci in a small population which leads to loss of uncommon allelic combinations, which increases the extent of LD in the population[20]. On the other hand, genetic drift brings new allele frequencies amongst populations, for instance, about 80% of the allele frequency has been lost because of domestication in maize[31], approximately 40-50% in sunflower crop[32] and 10-20% in rice[33].

Visual representation and statistical importance of linkage disequilibrium

To visualize the significant linkage disequilibrium, two methods are generally employed in the genomic regions of haplotypes. First, linkage disequilibrium scatter plots are considered to quantify the LD decay rate with genetic distance[27]. An average of the genome-wide decay of linkage disequilibrium can be evaluated by plotting a graph of linkage disequilibrium data set of the whole genome against the genetic or physical distance[28]. These dispersed plots are helpful to estimate the genetic distance at a significant LD threshold D′, 0.5 or r^2 0.1. Second, LD heat maps are also very significant, otherwise called disequilibrium matrices[20]. These are used to visualize LD between the polymorphic sites present within genes and the whole genome[20]. Linkage disequilibrium heat maps are statistical plots generally coded with color and triangular where the diagonals entitle the order of loci and the pixels with color represent the level of significant LD, which are calculated by D′ or r^2 values[28]. These graphical representations help to determine the marker-trait association to identify the QTLs in the map. PowerMarker[34] and TASSEL[35] are the most used computer software programme used for this approach.

Approaches and computer software packages used in association mapping studies

In a population structure, non-random mating leads to a change in allelic frequency[36], which is a major restriction in association mapping studies. Different methodologies are followed to minimize the false positive association and analyze the marker-trait association using the concept of LD (Table 2). In this table, the approaches and models which are generally performed in association mapping analysis like MAGIC, CC, TDT, GC, SA, GLM, MLM, CMLM, ECMLM, SUPER, MLMM and Farm CPU are presented. Several software packages are available for association mapping studies in plants. TASSEL (Trait analysis by association, evolution and linkage) is a commonly used software package for LD mapping population structure analysis, graphical representation of LD and finding out the genetic distance in plants[35]. TASSEL can process simple sequence repeats (SSR) and single nucleotide polymorphism (SNP) markers[35].

GenStat[47] is alternative software that generally offers traditional statistical analysis methods. Some useful computer software programs used for marker-trait association studies are summarized in Table 3

Table 2 — Analytical approaches and models generally used in association mapping analysis in plants

Approach	Description
MAGIC (Multiparent Advanced Generation Intercross)	Capable to capture major genetic variation in gene pool[37]
CC (Case-control)	Classical methodology used for population stratification of significant association at a particular loci[38]
TDT (Transmission Disequilibrium Test)	Linkage can be detected with a controls on false positive result[39].
GC (Genomic Control)	To tackle the problem of spurious association, this is used to find confounding effects of a population. It is having improved methodology involving additional markers which have neutral effect on the trait of interest in the study[40].
SA (Structured association)	Structure Association involves searching of closely related cluster for better dealing with structure population[41].
GLM (General Linear Model)	For the reduction in false positive results which arises because of population structure[15]
MLM (Mixed linear model)	For the reduction in false positive result which arises from family relatedness and population structure[15]
CMLM (Compressed MLM)	Single locus analysis, having better statistical power than other MLM[42]
ECMLM (Enriched CMLM)	Single locus analysis, to choose the best combination among grouping algorithms and kinship[43]
SUPER (Settlement of MLM Under Progressively Exclusive Relationship)	Single locus analysis to derive kinship and associated genetic markers are being used in this model instead of other markers[44].
MLMM (Multiple Loci Mixed Linear Model)	Multi locus analysis by integrating both kinship matrix and cofactors performing better in regards to the recovery rate and detection power[45]
Farm CPU (Fixed and random model Circulating Probability Unification)	Multi-locus analysis, removes confusion between testing marker and kinship by integrating two different models viz., MLM and MLMM method[46]

Table 3 — List of software used in association mapping analysis and source

Software	Focal point of the software package	Online source reference link
BIMBAM	Candidate genes and quantitative traits	https://omictools.com/bimbam-tool
BLIMP	Allelic frequencies detection	https://stephenslab.uchicago.edu/software.html
eQtlBma	Quantitative trait loci for gene expression levels	https://bio.tools/eQtlBma
BRIdGE	Evidence of gene and environment interaction	https://stephenslab.uchicago.edu/software_pages/bridge/index.html
fastPHASE	Estimates the missing genotypes	https://www.mybiosoftware.com/fastphase-1-4-0/
GEMMA	Genome-wide association studies	http://stephenslab.uchicago.edu/software.html
mvBIMBAM	Multivariate association analysis of related phenotypes	http://stephenslab.uchicago.edu/software.htm
EMMA	Population structure and genetic relatedness analysis	http://mouse.cs.ucla.edu/emma/index.html
STRUCTURE	Population structure and genetic relatedness analysis	http://pritch.bsd.uchicago.edu/software.html
BAPS	Population structure and genetic relatedness analysis	http://web.abo.fi/fak/mnf/mate/jc/software/baps.html
mStruct	Population structure and genetic relatedness analysis	http://www.cs.cmu.edu/suyash/mstruct.html
SPAGeDi	Characterise the individuals or mapping populations	http://www.ulb.oc.be/science/ecoevol/spagedi.html
EINGENSTRAT	Principal component analysis	http:/genepath.med.harvard.edu/reich/softwatware.html
LDheatmap	Linkage disequilibrium analysis	http://www.jstatsoft.org/v16/c03
MIDAS	Linkage disequilibrium analysis	http://www.genes.org.uk/software/midas/
LDhat	Linkage disequilibrium analysis	http://www.stats.ox.ac.uk/mcvean/ldhat.html
Arlequin	Linkage disequilibrium analysis	http://cmpg.unibe.ch/software/arlequin35/
Haploview	Haplotype analysis and linkage disequilibrium analysis	http://www.broad.mit.edu/mpg/haploview/
GGT	Haplotype analysis and linkage disequilibrium analysis	http://www.plantbreeding.wur.nl/UK/software_ggt.html
SVS	Haplotype analysis and linkage disequilibrium analysis	http://www.goldenhelix.com
TASSEL	Haplotype analysis and linkage disequilibrium analysis	http://www.maizegenetics.net
PLINK	Haplotype analysis and linkage disequilibrium analysis	http://pngu.mgh.harvard.edu/purcell/plink/

with their sources. This summary provides the information regarding the software packages such as BIMBAM (software for Bayesian IMputation- Based Association Mapping), BLIMP (software for Best Linear IM Putation), BRIdGE (Bayesian Regression for Identifying Gene-Environment Interactions),

eQtlBma: (software to detect eQTLs by Bayesian Model Averaging), fast PHASE (software for haplotype reconstruction, and estimating missing genotypes from population data), GEMMA (Genome-wide Efficient Mixed Model Association), mvBIMBAM (software for genetic association analysis of multiple related phenotypes, EMMA (Efficient Mixed-Model Association) and SPAGeDi (Spatial Pattern Analysis of Genetic Diversity), etc.

Marker and trait association studies in crop plants

Linkage mapping and linkage disequilibrium mapping are complementary. Large-scale genome-wide association studies in plants have already been completed and more are in progress. The technological trouble of successfully genotyping a million or greater SNPs has recently been solved[29]. Association analysis will now not be restrained to the identity of QTL however will also offer extensive information on the genomic modifications for crop domestication and selection. Such information will translate into the superior design of breeding populations and germplasm collections.

It is challenging because linkage disequilibrium mapping regularly involves a relatively significant amount of accession phenotypic data collection with replications for the duration of multiple years and locations. One of the limitations of linkage disequilibrium mapping is that, it offers little perception into the mechanistic basis of linkage disequilibrium detected. This limitation happens because of the sturdy linkage disequilibrium resulting from mutations in preference to a close linkage among loci displaying linkage disequilibrium[19]. Consequently, it is expected that multiplied usage of nested association mapping will encompass a number of studies related to the phenotypic variants observed for agronomic and adaptive traits in crops. The first study of marker-trait association[48] was conducted in maize and rice crop plants using random amplified polymorphic DNA markers and protein markers.

At the same time, thirteen putative QTLs were also mapped in oat landraces using polymorphic restriction fragmented length polymorphism markers[49]. The map's resolution was found to be less as fewer markers were implemented and distributed across the whole genome. In the model plant Arabidopsis, most association mapping analysis was done to find out the concept of linkage disequilibrium and association mapping[36]. The first genome-wide association study in Arabidopsis was then reported for flowering time and biotic stress resistance genes[50]. In Table 4, some recent studies of association mapping in crop plants were summarized, including the type of molecular marker used and the number of accessions taken for the experiments. The trait of interest exploited in these experiments is also mentioned in this table with their respective references.

Several trait loci, the interaction between them, the environment and the relationship between QTL and environment affect the phenotypic variability of several complex traits of agricultural or evolutionary significance. The association mapping can be utilized to classify genes for quantitative variability of complex traits of agricultural and evolutionary significance is of great interest. The associations mapping dissects the genetic variation present in natural populations that can effectively overcome complex differences in traits to have single genes or individual nucleotides[36]. Conventional study of the linkage with experimental populations resulting from a biparental cross offers valuable information on traits that appear to be specific to the same or genetically similar populations. In contrast, the findings of association mapping are more applicable to several other germplasm bases. Association mapping may also be used for the short-term domestication of neglected crops and wild plants[36].

Traditionally, plant domestication and the production of viable varieties required decades of breeding, which is also the main reason why most breeding efforts have focused on further developing a relatively limited number of crops already domesticated several thousand years. The recent discovery of several primary genes for domestication and technological breakthroughs in implementing several genomic changes in plants simultaneously with association mapping allows wild species to be domesticated within a certain plant generation[70]. Association mapping has the ability to map QTLs[70,71] in the collection of breeding lines, landraces, or samples from the natural population[72]. For the development of improved traits and germplasm securing association mapping has immense potential.

Table 4 — Association mapping studies in crop plants

Crop	Number of accessions	Marker type	Number of markers	Traits of interest
Rice	170 accessions	SSR	84	Plant height, Panicle length, Glume pubescence, Length or width of the ratio of grain, leaf of pubescence[51]
Rice	523 accessions	SNP	5291	Tiller number, tiller angle, plant height, flag leaf length. Flag leaf width. Flag leaf length and width, flag leaf angle, panicle number[52]
Wheat	95 accessions	SSR	36	Kernel area, kernel length, kernel width, superior milling score, higher flour, yield, friability [53]
Wheat	100 winter varieties	DArT	5525	Plant height, grain yield, bioethanol production[54]
Wheat	242 genotype	SSR	250	Tolerant to Pre-harvest, Sprouting, tolerant to PHS[55]
Wheat	230 genotypes	SSR	250	Plant height, peduncle length, flag leaf length, day to maturity, spike length, grain protein content [56]
Maize	289 lines	SNP	56110	21 amino acids, 13 organic acids, 7 phenyl propanoids[57]
Maize	302 lines	SSR	89	Kernel, protein, starch, and moisture content[20]
Barley	318 accessions	DArT; SNPs	558 DArT 2878 SNPs	Spot blotch resistant[58]
Sugarcane	154 clones	AFLP; SSR	1068 AFLP, 141 SSR	Root rot, leaf scald, Fiji leaf gall, Smut disease[59]
Sorghum	377 lines	SSR	47	Flag leaf length and width, plant height, flowering time[60]
Arabidopsis	95 accessions	SNPs	2553	Flowering and pathogen resistant [40]
Pepper	74 lines	SSR; RAPD; SCAR	290 SSR; 30 RAPD; 9 SCAR	Plant height, no. of fruit per plant, ten fruit weight, total fruit weight, fruit length, fruit width, pericarp thickness[61]
Egg plant	191 accessions	SNPs	70	Fruit weight, fruit length, fruit diameter, fruit position, fruit shape, fruit curvature, fruit apex shape[62]
Tomato	174 accessions	SSR	182	Fruit nutritional quality traits[63,64]
Mung bean	466 accessions	SNP	22000	Study of seed coat colour[65]
Peanut	104 accessions	SSR	554	Study on seed related traits[66]
Field Pea	135 accessions	SNP	16877	Agronomic and Seed Quality Traits[67]
Mung bean	293 accessions	SNP	5041	Agronomical traits[68]
Bread wheat	82 accessions	SNP	35000	Phosphorus efficiency traits[69]

(SSR – Simple sequence repeats, SNP – Single nucleotide polymorphism, DArT – Diversity array technology, AFLP – Amplified fragmented length polymorphism, RAPD – Random amplified polymorphic DNA, SCAR – Sequence characterized amplified region, PHS – Pre-harvest sprouting)

Conclusion

Association mapping is a useful tool for exploiting the traditional landraces cultivated by farmers to enhance their genetic diversity panel. So, exploring the narrow genetic diversity to make it comprehensive for identifying novel stress resistance genes in the traditionally cultivated varieties grown by a farmer using the association mapping technique helps the plant breeders to develop new varieties with improved agronomic traits. Missing heritability, refinements in experimental design, statistical analysis, development of new algorithms for efficient detection of epistasis, population structure, accurate phenotyping, the discovery of rare alleles of genes/QTLs associated with the rare marker alleles, association studies and marker-assisted selection, mapping of QTLs by jointly using linkage and LD, and identification of QTNs (Quantitative trait nucleotides) in crop plant species are the future challenges to be explored for improvement in the plant breeding strategies by association mapping.

Conflict of Interest

The authors declare that they have no conflict of interest.

Authors' Contributions

The authors confirm contribution to the paper as follows: Conceptualization, JPS; investigation and writing the original draft, JPS, UM and DS; review and editing, KCS and ICM. All authors have read and agreed to the published version of the manuscript.

References

1   Meilleur B A & Hodgkin T, *In situ* conservation of crop wild relatives: status and trends, *Biodivers Conserv,* 13 (4) (2004) 663-684.

2   Ross-Ibarra J, Morrell P L & Gaut B S, Plant domestication, a unique opportunity to identify the genetic basis of adaptation, *Proc Natl Acad Sci*, 104 (2007) 8641-8648.

3   Yu J & Buckler E S, Genetic association mapping and genome organization of maize, *Curr Opin Biotechnol*, 17 (2006) 155-160.

4   Nuez F & Bermejo J H, Neglected crops: 1492 from a different perspective. Plant production and protection series (1994) 303-32.

5   Mackay T F, The genetic architecture of quantitative traits, *Annu Rev Genet*, 35 (2001) 303–339.

6   Collard B C, Jahufer M Z, Brouwer J B & Pang E C, An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts, *Euphytica*, 142 (1-2) (2005) 169-96.

7   Liu B H, Statistical Genomics: Linkage, Mapping and QTL Analysis, CRC Press, New York, NY, USA (1998).

8   Wu R, Ma C & Casella G (2007) Statistical Genetics of Quantitative Traits: Linkage, Maps and QTL, Springer, New York.

9   Flint-Garcia S A, Thornsberry J M & Buckler E S, Structure of linkage disequilibrium in plants, *Annu Rev Plant Biol*, 54 (2003) 357-74.

10   Buckler E S & Thornsberry J M, Plant molecular diversity and applications to genomics, *Curr Opin Plant Biol*, 5 (2002) 107-111.

11   Gordon D & Finch S J, Factors affecting statistical power in the detection of genetic association, *J Clin Invest*, 115 (2005) 1408-18.

12   Altshuler D, Daly M J & Lander E S, Genetic mapping in human disease, *Sci*, 322 (2008) 881-888.

13   Rafalski J A, Association genetics in crop improvement, *Curr Opin Plant Biol*, 13 (2010) 174–180.

14   Visscher P M, Sizing up human height variation, *Nat Genet,* 40 (2008) 489–490.

15   Yu J, Pressoir G, Briggs W H, Bi I V, Yamasaki M, *et al.*, A unified mixed-model method for association mapping that accounts for multiple levels of relatedness, *Nat Genet*, 38 (2006) 203-208.

16   Flint-Garcia S A, Thuillet A C, Yu J, Pressoir G, Romero S M, *et al.*, Maize association population: a high-resolution platform for quantitative trait locus dissection, *Plant J*, 44 (2005) 1054-64.

17   Zhu C, Gore M, Buckler E S & Yu J, Status and prospects of association mapping in plants, *Plant Genome*, 1 (2008) 5-20.

18   Chan E K, Rowe H C & Kliebenstein D J, Understanding the evolution of defense metabolites in *Arabidopsis thaliana* using genome-wide association mapping, *Genetics*, 185 (2010) 991-1007.

19   Gupta P K, Rustgi S & Kulwal P L, Linkage disequilibrium and association studies in higher plants: present status and future prospects, *Plant Mol Biol*, 57 (2005) 461-85.

20   Flint-Garcia S A, Thornsberry J M & Buckler IV E S, Structure of linkage disequilibrium in plants, *Annual Rev Plant Biol*, 54 (2003) 357-374.

21   Stich B & Melchinger A, An introduction to association mapping in plants, CAB Reviews: Perspectives in agriculture, *Vet Sci Nutr Nat Resour*, 5 (2010) 1-9.

22   Soto Cerda B J & Cloutier S, Association mapping in plant genomes, Genetic diversity in plants, (Intech open), 2012, p. 29-54.

23   Lewontin R C, The interaction of selection and linkage, I General considerations; heterotic models, *Genetics*, 49 (1964) 49.

24   Dawson E, Abecasis G R, Bumpstead S, Chen Y, Hunt S, *et al.,* A first-generation linkage disequilibrium map of human chromosome 22, *Nature*, 418 (2002) 544-548.

25   Garris A J, McCouch S R & S Kresovich S, Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the xa5 locus of rice (*Oryza sativa* L.), *Genetics,* 165 (2003) 759–769.

26   Malysheva-Otto L & Ganal M R, Analysis of molecular diversity, population structure and linkage disequilibrium in a worldwide survey of cultivated barley germplasm (*Hordeum vulgare* L.), *BMC Genet*, 7 (2006) 1-14.

27   Garris A J, Tai T H, Coburn J, Kresovich S & McCouch S, Genetic structure and diversity in *Oryza sativa* L., *Genetics*, 169 (2005) 1631-8.

28   Zhao, K, María J A, Sung K, Clare L, Chikako S, *et al.*, An Arabidopsis example of association mapping in structured samples, *PLoS Genet,* 3 (1) (2007) e4.

29   Slatkin M, Linkage disequilibrium: Understanding the evolutionary past and mapping the medical future, *Nat Rev Genet*, 9 (2008) 477-485.

30   Thornsberry J, Goodman M, Doebley J, Kresovich S, Nielsen D, *et al.*, Dwarf 8 polymorphisms associate with variation in flowering time, *Nat Genet*, 28 (2001) 286-289.

31   Wright S & Gaut B, Molecular population genetics and the search for adaptative evolution in plants, *Mol Biol Evol*, 22 (2005) 506-519.

32   Liu A & Burke J M, Patterns of nucleotide diversity in wild and cultivated sunflower, *Genetics*, 173 (2006) 321-30.

33   Zhu Q, Zheng X, Luo J, Gaut B S & Ge S, Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: severe bottleneck during domestication of rice, *Mol Biol Evol*, 24 (2007) 875-888.

34   Liu K & Muse S V, Power Marker: an integrated analysis environment for genetic marker analysis, *Bioinformatics*, 21 (2005) 2128-2129.

35   Bradbury P J, Zhang Z, Kroon D E, Casstevens T M, Ramdoss Y, *et al.,* TASSEL: software for association mapping of complex traits in diverse samples, *Bioinformatics*, 23 (2007) 2633-2635.

36   Ersoz E S, Yu J & Buckler E S, Applications of linkage disequilibrium and association mapping in crop plants, In: *Genomics-assisted crop improvement*, Springer, Dordrecht, (2007) p. 97-119.

37   Mackay I & Powell W, Methods for linkage disequilibrium mapping in crops, *Trends Plant Sci*, 12 (2007) 57-63.

38   Morton N E & Collins A, Tests and estimates of allelic association in complex inheritance, *Proc Natl Acad Sci*, 95 (1998) 11389-11393.

39   Spielman R S & Ewens W J, The TDT and other family-based tests for linkage disequilibrium and association, *Am J Hum Genet*, 59 (1996) 983.

40   Devlin B & Roeder K, Genomic control for association studies, *Biometrics*, 55 (1999) 997-1004.

41   Pritchard J K, Stephens M, Rosenberg N A & Donnelly P, Association mapping in structured populations, *Am J Hum Genet,* 67 (2000) 170-181.

42   Zhang Z, Ersoz E, Lai C Q, Todhunter R J, Tiwari H K, *et al*., Mixed linear model approach adapted for genome-wide association studies, *Nat Genet,* 42 (2010) 355.

43   Li M, Liu X, Bradbury P, Yu J, Zhang Y M, *et al*., Enrichment of statistical power for genome-wide association studies, *BMC Biol*, 12 (2014) 73.

44   Wang Q, Tian F, Pan Y, Buckler E S & Zhang Z, A SUPER powerful method for genome wide association study, *PloS one*, 9 (2014) e107684.

45   Segura V, Vilhjálmsson B J, Platt A, Korte A, Seren Ü, *et al*., An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations, *Nat Genet*, 44 (2012) 825.

46   Liu X, Huang M, Fan B, Buckler E S & Zhang Z, Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies, *PLoS Genet,* (2016) e1005767.

47   Emebiri LC, Matassa V & Moody D B, GENSTAT programs for performing Muir's alternative partitioning of genotype-by-environment interaction, J *Hered,* 96 (2005) 78-9.

48   Virk P, Ford-Lloyd B, Jackson M, Pooni H, Clemeno T, *et al*., Predicting quantitative variation within rice germplasm using molecular markers, *Heredity,* 76 (1996) 296-304.

49   Beer S C, Siripoonwiwat W, O'donoughue L S, Souza E, Matthews D, *et al*., Associations between molecular markers and quantitative traits in an oat germplasm pool: can we infer linkages, *J Agri Genom*, 3 (1997) 1-6.

50   Aranzan M, Kim S, Zhao K, Bakker E, Horton M, *et al*., Genome-wide association mapping in Arabidopsis identifies previously known flowering time and pathogen resistance genes, *PLoS Genet,* 1 (2005) 0531-0539.

51   Wen W, Mei H, Feng, Yu S, Huang Z, *et al*., Population structure and association mapping on chromosome 7 using a diverse panel of Chinese germplasm of rice (*Oryza sativa* L.), *Theor Appl Genet* 119 (2009) 459-470.

52   Lu Q, Zhang M, Niu X, Wang S, Xu Q, *et al*., Genetic variation and association mapping for 12 agronomic traits in indica rice, *BMC Genomics*, 16 (2015) 1067.

53   Breseghello F & Sorrells M E, Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars, *Genetics*, 172 (2006) 1165-1177.

54   Bellucci A, Torp A M, Bruun S, Magid J, Andersen S B, *et al*., Association mapping in scandinavian winter wheat for yield, plant height and traits important for second-generation bioethanol production, *Front Plant Sci*, 6 (2015) 1046.

55   Jaiswal V, Mir R R, Mohan A, Balyan H S & Gupta P K, Association mapping for pre-harvest sprouting tolerance in common wheat (*Triticum aestivum* L.), *Euphytica*, 188 (2012) 89-102.

56   Jaiswal V, Gahlaut V, Meher P K, Mir R R, Jaiswal J P, *et al*., Genome wide single locus single trait, multi-locus and multi-trait association mapping for some important agronomic traits in common wheat (*T. aestivum* L.), *PloS One*, 11 (2016) e0159343.

57   Riedelsheimer C, Lisec J, Czedik-Eysenberg A, Sulpice R, Flis A, *et al.,* Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize, *Proc Natl Acad Sci,* 109 (2012) 8872-8877.

58   Roy J K, Smith K P, Muehlbauer G J, Chao S, Close T J, *et al.,* Association mapping of spot blotch resistance in wild barley, *Mol Breed*, 26 (2010) 243-256.

59   Wei X, Jackson P A, McIntyre C L, Aitken K S & Croft B, Associations between DNA markers and resistance to diseases in sugarcane and effects of population substructure, *Theor Appl Genet*, 114 (2006) 155-164

60   Casa A M, Pressoir G, Brown P J, Mitchell S E, Rooney W L, *et al.,* Community resources and strategies for association mapping in sorghum, *Crop Sci*, 48 (2008) 30-40.

61   Dwivedi N, Kumar R, Paliwal R, Kumar U, Kumar S, *et al.,* QTL mapping for important horticultural traits in pepper (Capsicum annuum L.), *J Plant Biochem Biotechnol*, 24 (2015) 154-60.

62   Portis E, Cericola F, Barchi L, Toppino L, Acciarri N, *et al.,* Association mapping for fruit, plant and leaf morphology traits in eggplant, *PLoS One*, 10 (2015) e0135200.

63   Zhang J, Zhao J, Xu Y, Liang J, Chang P, *et al.,* Genome-wide association mapping for tomato volatiles positively contributing to tomato flavor, *Front Plant Sci*, 6 (2015) 1042.

64   Stich B& Melchinger A, An introduction to association mapping in plants, CAB Reviews: Perspectives in agriculture, *Vet Sci Nutr Nat Resour*, 5 (2010) 1-9.

65   Noble T J, Tao Y, Mace S, Williams B, Jordan D R, *et al.,* Characterization of linkage disequilibrium and population structure in a mungbean diversity panel, *Front Plant Sci*, 8 (2018) 2102.

66   Zhao J, Huang L, Ren X, Pandey M K, *et al.,* Genetic variation and association mapping of seed-related traits in cultivated peanut (*Arachis hypogaea* L.) using single-locus simple sequence repeat markers. *Front Plant Sci*, 8 (2017) 2105.

67   Gali K K, Sackville A, Tafesse E G, Lachagari V B, McPhee K, *et al.,* Genome-wide association mapping for agronomic and seed quality traits of field pea (*Pisum sativum* L.), *Front Plant Sci*, 10 (2019) 1538.

68   Sokolkova A, Burlyaeva M, Valiannikova T, Vishnyakova M, Schafleitner R, *et al.,* Genome-wide association study in accessions of the mini-core collection of mungbean (*Vigna radiata*) from the World Vegetable Gene Bank (Taiwan), *BMC Plant Biol*, 20 (2020) 1-9.

69   Soumya P R, Burridge A J, Singh N, Batra R, Pandey R, *et al.,* Population structure and genome-wide association studies in bread wheat for phosphorus efficiency traits using 35 K Wheat Breeder's Affymetrix array, *Sci Rep*, 11 (2021) 1-17.

70   Sahoo J P, Sharma V, Verma R K, Chetia S K, Baruah A R, *et al.,* Linkage analysis for drought tolerance in kharif rice of Assam using microsatellite markers, *Indian J Tradit Know*, 18 (2019) 371-375.

71   Sahoo J P, Behera L, Sharma S S, Praveena J, Nayak S K, *et al.,* Omics studies and systems biology perspective towards abiotic stress response in plants, *American J Plant Sci*, 11 (2020) 2172.

72   Manjunathagowda D C, Mb A, Jayaswalla K, Rc V, Aa K, *et al.,* Variability and genetic diversity among selfed lines (S1) of onion (*Allium cepa* L.), *Indian J Tradit Know*, 20 (2021) 563-568.