



Association Mapping Studies in Crop Plants: Concept and Applications

B. Kalyana Babu*, R. Arun Kumar and Lakshmi Kant

ICAR-Vivekananda Parvatiya Krishi Anusanthan Sansthan (VPKAS), Almora, Uttarakhand

*Email of corresponding author: kalyan_biotek@yahoo.co.in

Association mapping (AM) refers to the significant association of a marker locus with a phenotype trait whereas, linkage disequilibrium (LD) refers to the non random association between two markers or two genes/Quantitative trait loci (QTLs). The terms LD and AM have often been used interchangeably in the literature. In QTL mapping, LD is generated by the mating design whereas, in AM, LD is a reflection of the germplasm collections under study. AM has the potential of high resolution in localizing a QTL controlling a trait of interest is the primary advantage. Recently a lot of progress has been made in the application of AM in several crops. A variety of software packages are available for AM, and it can be inferred that LD studies are computationally demanding.

Introduction

Over the past decade, marker-assisted selection (MAS) has become a standard tool in breeding programs (Collard *et al.*, 2005). As a first step to marker-assisted selection (MAS), marker loci need to be identified that are linked to QTL. The objective of genetic mapping is to identify simply inherited markers in close proximity to genetic factors affecting quantitative traits (Quantitative trait loci, or QTL). The QTL mapping has largely been accomplished using linkage mapping of QTL (Singh *et al.*, 2012). However, the focus is now trending towards the use of association mapping which was initially applied in human disease genetics. Both the approaches had its own advantages and disadvantages. In the present article we are describing the details of association mapping, differences with linkage mapping, association mapping in plants and software used for the association mapping studies.

Association Mapping (AM)/ Linkage Disequilibrium (LD)

The terms LD and AM have often been used interchangeably in the literature. However, they present subtle differences. Association mapping refers to the significant association of a marker locus with a phenotype whereas while LD refers to the non random association between two markers or two genes/QTLs (Guptha *et al.*, 2005), AM is actually an application of LD. Both QTL and AM approaches are, therefore, based on LD between molecular markers and functional loci. In QTL mapping, LD is generated by the mating design whereas in AM, LD is a reflection of the germplasm collections under study (Stich *et al.*, 2010). The concept of LD was first described by Jennings in 1917, and its quantification (D) was developed by Lewontin in 1964. For linkage studies, suitable crosses, sometimes limited by low polymorphism or small population size, are required. Only two alleles per locus and few recombination events are considered to estimate the genetic distances between marker loci and to identify the causative genomic regions for quantitative trait loci (QTL). To circumvent these limitations, linkage disequilibrium (LD) mapping or association mapping (AM) has been used in human genetic studies and then further being effectively used in plants.

Advantages of Association Mapping (AM)

1. The primary advantage is AM has the potential high resolution in localizing a QTL controlling a trait of interest.
2. A single polymorphism within a gene that is responsible for phenotypic differences can be identified.
3. Involves searching for genotype-phenotype correlations among unrelated individuals
4. High resolution is accounted for by the historical recombinations accumulated in natural populations and collections of landraces, breeding materials and varieties.
5. Time saving in establishing a marker-trait association and its application in a breeding program (Flint-Garcia *et al.*, 2010).
6. Identify more and superior alleles and provide detailed marker data in a large number of lines which could be of immediate application in breeding.
7. AM uses breeding populations including diverse and important materials in which the most relevant genes should be segregating.
8. Sample size and structure do not need to be as large as for linkage studies to obtain similar power of detection.

Disadvantages

1. Biological and evolutionary factors affect LD distribution and mapping resolution.
2. The strength of AM relies on the analysis of common variants, which explain at most 5%-10% of the heritable component.
3. LD mapping often requires a large number of markers for genotyping in GWAS.
4. When statistical methods to correct for population structure are applied, the differences between subpopulations are disregarded when searching for marker-trait associations.
5. The power of AM to detect an association is influenced by allele frequency distribution at the functional polymorphism level.

AM Studies in Plants

Rice (*Oryza sativa* L.) is a staple food for more than half of the world population. Rice landraces have evolved from their wild progenitor under natural and human selection, leading to the maintenance of high genetic diversity. Agrama *et al* (2007) conducted Association mapping of yield and its components in rice cultivars. Ninety-two rice germplasm accessions recently introduced from seven geographic regions of Africa, Asia, and Latin America, and eleven US cultivars, included as checks, were evaluated for yield and kernel characteristics, and genotyped with 123 SSR markers. Huang *et al* (2010) have identified ~3.6 million SNPs by sequencing 517 rice landraces and constructed a high-density haplotype map of the rice genome using a novel data-imputation method. They performed genome-wide association studies (GWAS) for 14 agronomic traits in the population of *Oryza sativa indica* subspecies. Cockram *et al* (2008) by taking barley as a typical inbreeding crop, and seasonal growth habit as a major partitioning phenotype, developed an association mapping approach which successfully identifies *VRN-H1* and *VRN H2*, the underlying loci largely responsible for this agronomic division. They successfully demonstrated association mapping of *VRN-H1* and *VRN-H2* in a collection of European barley germplasm, despite the prominent role of these genes in defining germplasm structure.

Cockram *et al* (2010) has conducted Genome-wide association mapping to candidate polymorphism resolution in the un-sequenced barley genome. Using 32 phenotypes in the inbreeding crop barley, they reported GWA mapping of 15 morphological traits across 500 cultivars genotyped with 1,536 SNPs. In contrast to the majority of human GWA studies, they observed high levels of linkage disequilibrium within and between chromosomes. Wang *et al* (2012) studied the effect of population size and unbalanced data sets on QTL detection using genome-wide association mapping in barley breeding germplasm. They examined 766

spring barley breeding lines grown in balanced trials and detected three major QTL for heading date.

Tian *et al* (2011) used Genome-wide association study of leaf architecture in the maize nested association mapping (NAM) population. GWAS in diverse maize is challenging, as linkage disequilibrium decays within 2,000 bp. NAM design to dissect the genetic architecture of upper leaf angle, leaf length and width. Using joint stepwise regression, they identified 30–36 quantitative trait loci (QTLs) for the three leaf traits, explaining 74.8%–80.3% of the phenotypic variation and >83% of genetic variance.

A genome-wide association study of wheat was reported, in which a large number of diversity array technology markers was used to genotype a winter wheat core collection of 96 accessions (Newmann *et al.*, 2011). Twenty agronomic traits were measured in field trials conducted over up to eight growing seasons and found in total, 385 significant marker–trait associations.

Association Mapping for Abiotic Stress Tolerance

Nutrient deficiencies, toxicities, cold, drought, and heat stress are widespread problems in agriculture areas. In order to survive and produce under highly variable environments, plants have developed multiple strategies and mechanisms of tolerance. The majority of traits associated with abiotic stress tolerance in plants are quantitative with complex phenotype and genetic control. Association or linkage disequilibrium (LD) mapping represents an alternative approach to identifying genes or genomic regions associated with quantitative phenotypic variation. Similar to QTL mapping, association mapping exploits natural diversity and recombination within a population to correlate polymorphisms with measurable phenotypic variations. However, in contrast to linkage or QTL mapping, which depends on recombination events generated over a fixed number of generations following a bi-parental cross, association mapping exploits larger number of historical recombination events in a population or diverse set of lines over the course of evolution. Using of association mapping along with QTL mapping is most productive in crop improvement for abiotic stress management.

Software used for Association Mapping

A variety of software packages are available for AM, and it can be inferred from the previous sections that LD studies are computationally demanding. More powerful programs are constantly under development. The following table illustrates the software being widely used for association mapping.

Table 1: The Software Widely used for Structure Analysis and Association Mapping Studies

Software	Focus	Website
STRUCTURE 2.3	Population structure	http://pritch.bsd.uchicago.edu/software.html
Arlequin 3.5	Genetic analysis and LD	http://cmpg.unibe.ch/software/arlequin35/
Haploview 4.2	Haplotype analysis and LD	http://www.broad.mit.edu/mpg/haploview/
TASSEL	LD and AM	http://www.maizegenetics.net
JMP genomics	LD and structured AM	http://www.jmp.com/software/genomics

Future Perspectives

Genome wide association study (GWAS) will not be limited to the identification of QTL but will also provide in depth understanding of the genomic changes that have shaped crop plants as a consequence of domestication and selection. Design and implementation of genotyping assays is no longer time-consuming or expensive. To fully exploit and benefit from the large amount of achievable genotyping data, care must be given to proper and powerful experimental design.

Conclusion

Association mapping has several advantages like it has Potential of high resolution in localizing a QTL controlling a trait of interest and a single polymorphism within a gene that is responsible for phenotypic differences can be identified. This association mapping will be very useful in marker assisted breeding studies.

References

- Agrama HA and Eizenga W Yan. 2007. Association mapping of yield and its components in rice Cultivars. *Mol Breeding.*, 19: 341–356.
- Cockram J, White J, Leigh FJ, Lea VJ, Chiapparino E, Laurie DA, Mackay IJ, Powell W and Sullivan DMO (2008). Association mapping of partitioning loci in barley *BMC Genetics.*, 9:16.
- Cockram J, White J, Zuluaga DL, Smith D, Comadran J, Macaulayb M, Luoc Z, Kearsey MJ, Werner P, Harrap D, Tapsell C, Liub H, Hedley PE, Steine N, Schultee D, Steuernagele B, Marshallb DF, Thomasb WTB, Ramsayb L, Mackaya I, Baldingf DJ, Consortium TA, Waughb R, and Sullivana DMO. 2010. (Genome-wide association mapping to candidate polymorphism resolution in the un-sequenced barley genome. *Proc. Natl. Acad. Sci.*, 107: 21611-21616.
- Collard BCY, Jahufer MZZ, Brouwer JB and Pang ECK. 2005. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. *Euphytica*, 142: 169–196.
- Flint-Garcia S, Thornsberry J and Buckler E. 2010. Structure of linkage disequilibrium in plants. *Annual Review Plant Biology.*, 54: 357-374.
- Gupta P, Rustgi S and Kulwal P. 2005. Linkage disequilibrium and association studies in higher plants: present status and future prospects. *Plant Molecular Biology.* 57, 461-485.
- Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Li W, Lin Z, Buckler ES, Qian Q, Zhang QF, Li J, Han B. 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genetics.*, 42: 961-967.
- Neumann K, Kobiljski B, Dencic S, Varshney RK, Borner A. 2011. Genome-wide association mapping: a case study in bread wheat (*Triticum aestivum* L.). *Mol Breeding*, 27:37–58.
- Singh UM, Babu BK and Anil Kumar. 2012. Quantitative trait loci (QTL) mapping: concepts and applications. *Biotech Today.*, 2: 37-38.
- Stich B and Melchinger A. 2010. An introduction to association mapping in plants. *CABReviews: Perspectives in agriculture, Veterinary Science, Nutrition and Natural Resources*, 5: 1-9.
- Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-G S, Rocheford TR, McMullen MD, Holland JB and Buckler ES. 2011. Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat Genet* 43: 159–162.
- Wang H, Smith KP, Combs E, Blake T, Horsley RD, Muehlbauer GJ. 2012. Effect of population size and unbalanced data sets on QTL detection using genome-wide association mapping in barley breeding germplasm. *Theor Appl Genet.*, 124:111–124.