NEW ERA AGRICULTURE MAGAZINE

Association Mapping: a powerful tool for crop improvement

Javed*, Jainendra Pratap, Aman Srivastava, Anamish Tyagi, Sonali Srivastava, Raja Bhaiya

Introduction:

Association mapping is a high-resolution method for mapping quantitative trait loci based on principle of linkage disequilibrium that holds a great promise for the dissection of complex genetic traits. It is a powerful tool for the dissection of complex agronomic traits and for the identification of alleles that can contribute to the enhancement of a target trait. The power of association studies is determined by the size of the experimental population, the magnitude of the target allele effect, the density of markers used, and the rate of LD decay between marker and target allele as well as errors in phenotyping and genotyping data and the desired resultant statistical significance level. Association mapping is a very efficient and effective method for confirming candidate genes or for identifying new genes. It is now being increasingly used in a wide range of plants, where it appears to be more powerful than in humans or animals. Though association mapping is widely used, it has a lower power to detect rare alleles in a population, even those with large effects, than linkage mapping.

Association mapping is a useful alternative to standard QTL mapping approaches which involves the correlation of molecular polymorphisms with phenotypic variation in a diverse assemblage of individuals. The comparatively high-resolution provided by association mapping is dependent upon the structure of linkage disequilibrium (LD) across the genome. Association studies can be divided into two broad categories:

(i) Candidate gene association mapping

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

(ii) **Genome Wide Association mapping**

Here, genetic variation is explored within the whole genome, aiming to find signals of association with the complex trait. GWA mapping is a promising method to identify novel loci involved in complex phenotypic traits. However, GWA mapping should not be regarded as a replacement of traditional QTL mapping.

Candidate Gene Strategy

The candidate gene method of association analysis is a hypothesis-driven

Javed, Jainendra Pratap, Aman Srivastava, Anamish Tyagi, Sonali Srivastava, Raja Bhaiya*

Research Scholar, Dep. of Genetics and Plant Breeding, Acharya Narendra Deva University of Agriculture and Technology, Kumarganj, Ayodhya, U.P.

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

Genome Wide Association Mapping

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

Quantitative trait locus (QTL) mapping and association mapping are the most commonly used tools for dissecting the genetic basis of phenotypic trait variation. In QTL mapping only a limited number of recombination events that have occurred within families and pedigrees can be studied, whereas with association mapping the recombination events that have accumulated over thousands of generations can be exploite. Since the 1980s, QTL mapping has been used

most frequently, but association mapping is a promising alternative method for dissecting complex traits. Increased mapping resolution, reduced research time and larger allele numbers have been put forward as main advantages over traditional QTL mapping.

General Procedure of Association Mapping

The general procedure for genome-wide association mapping in plants is briefly outlined here based on Abdurakhmonov and Abdukarimov (2008).

1. Association Mapping Population

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

2. Phenotyping

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

3. Genotyping for Population Structure Analysis

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

4. Structure and Kinship Analysis

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

5. Genotyping for LD Analysis

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

6. AM and LD Analyses

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

Statistical Approaches Uses for Association Mapping

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

LD between a single marker and a QTL can be measured by regression analysis, where the data on the trait is regressed on the individual marker genotypes, so that significant regressions will identify the markers associated with the phenotype. Nowadays several software uses to assess the association of marker loci with traits. The most commonly used statistics include logistic regression with the possibility of structured associations implemented in TASSEL General Linear Model (Yu and Buckler, 2006, TASSEL: http://www.maizegenetics.net), a multiple regression model combined with the estimates for the false discovery rate suggested by Kraakman et al., (2006), and a unified mixed-model approach described by Yu et al., (2006) and implemented in TASSEL Mixed Linear Model or in SAS v9.1.2 (Ehrenreichet al., 2007).

Advantage of Association Mapping

Association mapping is a valuable tool for the detection of novel genes or QTLs of important agronomic characteristics. The extensive application of this approach in crop plants is expected in the long term as a result of establishment of the novel high-throughput genotyping and sequencing technologies. Gene-based markers are more accurate than linked markers for the prediction of phenotype, since the marker–trait association do not lost during segregation in the course of recurrent breeding selection cycles. Results from association analysis can be used to predict the best haplotype across one or multiple genes for optimum expression of the target trait.

Genome-wide association studies are currently exploited for mapping of disease genes in human genetics. In crop plants, the potential of exploiting LD to detect marker-trait associations was recently investigated for maize, wheat, barley, sorghum, ryegrass, soybean and rice. Association studies based on correlations between alleles at different sites or LD can provide high resolution for the identification of genes that contribute to phenotypic variation in natural populations. This approach has a potential to identify a single polymorphism within a gene that is responsible for the difference in phenotype. In addition, many plant species have high levels of diversity for which association approaches are well suited to evaluate the numerous alleles available. LD plays a central role in association analysis. The distance over which LD persists will determine the number and density of markers, and experimental design needed to perform an association analysis.

Linkage Disequilibrium

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

Summaries

Association mapping offers great potential to enhance crop genetic improvement. This is strengthened by the use of high throughput and cost effective next generation sequencing techniques that will enable GWA studies to become a popular and routine approach. However, association mapping remains complementary as replacement for linkage mapping and other gene identification and validation techniques. Moreover, the contrast between the large numbers of variants with small effects identified by GWA studies versus the small number of genomic regions with large effects

identified by linkage mapping remains a challenge to our current understanding of the genetic architecture of complex traits. Although, for practical applications, the integration of linkage mapping and association mapping approaches offers substantial opportunity to resolve the individual constraints of each approach while synergizing their respective strengths. Population structure remains a big limitation for association studies that requires careful choice of germplasm and the development of advanced statistical approaches.

References:

1. Singh, B.D., Singh, A.K. 2016. Marker assisted plant breeding: principles and practices, Springer
2. Kushwaha, U. K. S., Mangal, V., Bairwa, A. K., Adhikari, S., Ahmed, T., Bhat, P., ... & Singh, N. K. (2017). Association mapping, principles and techniques. *J. Biol. Environ. Eng*, *2*(1), 1-9.
7. 3.Dooner, H.K., and L. He 2008. Maize genome structure variation: Inter-play between retrotransposon polymorphisms and genic recombina-tion. Plant Cell 20:249–258.
3. Ehrenreich I.M., Stafford P.A. and Purugganan M.D. 2007. The ge-netic architecture of shoot branching in Arabidopsis thaliana: A comparative assessment of candidate gene associations vs. quantitative trait locus mapping. Genetics 176:1223-1236.
4. Oraguzie, N.C., Rikkerink E.H.A., Gardiner S.E. and Silva H.N. de 2007. Association Mapping in Plants. Springer, Tokio and New York, 277 pp.