



ISSN: 0975-833X

## RESEARCH ARTICLE

### ASSOCIATION MAPPING TO DISSECT AND EXPLOIT EXISTING NATURAL VARIATION FOR CROP IMPROVEMENT

<sup>1,\*</sup>Veerasha, B. A., <sup>1</sup>Rudra Naik, V., <sup>1</sup>Suma S. Biradar, <sup>1</sup>Desai, S. A. and <sup>2</sup>Chetti, M. B.

<sup>1</sup>Department of Genetics and Plant Breeding, University of Agricultural Sciences, Dharwad -580005, India

<sup>2</sup>Assistant Director General (HRD), Indian Council of Agricultural Research, New Delhi, India

#### ARTICLE INFO

##### Article History:

Received 19<sup>th</sup> January, 2015

Received in revised form

17<sup>th</sup> February, 2015

Accepted 20<sup>th</sup> March, 2015

Published online 28<sup>th</sup> April, 2015

##### Key words:

Association Mapping,

Natural Variation,

Complex Traits,

Genotyping,

Phenotyping,

Crop Improvement

#### ABSTRACT

The germplasm is often exploited to develop improved crop varieties for changing needs and environments. The vast amount of genetic variation present in the germplasm can be exploited to the best when the economically important traits have been mapped with molecular markers using specially designed mapping populations. Most of yield contributing traits are controlled by many loci and their molecular characterization and genetic mapping is called quantitative trait loci mapping (QTL-mapping). However, it is always not possible or convenient to construct such mapping populations in all crop plants. For example, in many tree species, pure lines or inbred lines, it is difficult if not impossible to construct mapping populations. Under such situations and in order to achieve higher level of resolution and precision in QTL analysis the association mapping technique is used. In the present review, we define and distinguished between linkage analysis based mapping association mapping and linkage disequilibrium (LD) mapping, and then described the technique of association mapping used to dissect the genetic basis of complex traits for crop improvement. This paper will be a key reference for the beginners and research scholars who are involved in association mapping in crop plants.

Copyright © 2015 Veerasha et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

#### INTRODUCTION

Dissecting the genetic basis of economically important complex traits is a major goal of plant breeding. The term "complex trait" refers to any trait that does not show classic Mendelian recessive or dominant inheritance attributable to a single gene locus. Most of the important traits in plant breeding viz., yield, height, drought resistance, disease resistance in many crop species are quantitative or complex traits. (Holeyachi *et al.*, 2013) The genetic variation of a quantitative trait is assumed to be controlled by the collective effects of many genes. These are known as quantitative trait loci (QTLs). A QTL is defined as a region of the genome that is associated with an effect of a quantitative trait. So a QTL can be a single gene, or it may be a cluster of linked genes that affect the traits (Veerasha *et al.*, 2015). QTL mapping is extremely useful in detecting various regions of genome that affect the expression of complex traits in a large number of species (Mauricio 2001). However, it suffers from number of limitations. Firstly, as only two parents are used to construct QTL mapping populations, allelic variation in each cross is restricted.

Secondly, recombination events per chromosome are usually less since early generation crosses are used, thus, limiting the resolution of the genetic map. Thirdly, the development of mapping populations is either not possible or is very time consuming in some crop species and most forest trees, which have resulted in prevention or slowing down of any progress in dissecting the genetic basis of complex traits using QTL mapping. However, the focus is now turning towards the use of association or linkage disequilibrium mapping, which has been hailed as a more efficient way of determining genetic basis of complex traits.

#### What is association mapping?

The terms linkage disequilibrium and association mapping have often been used interchangeably in literature but a distinction should be made. Association mapping refers to significant association of a molecular marker with a phenotypic trait, while, LD refers to non-random association between two markers or two genes/QTLs or between a gene/QTL and a marker locus. Thus, association mapping is actually one of the several uses of LD. In statistical sense, association refers to covariance of a marker polymorphism and a trait of interest, while LD represents covariance of polymorphisms

\*Corresponding author: Veerasha, B. A.

Department of Genetics and Plant Breeding, University of Agricultural Sciences, Dharwad -580005, India.

exhibited by two molecular markers/genes. Association mapping identifies QTLs by examining the marker-trait associations that can be attributed to the strength of linkage disequilibrium between markers and functional polymorphisms across a set of diverse germplasm lines (Gupta *et al.*, 2005).

### Why association mapping?

The phenotypic variation of many complex traits of agriculturally important crops is influenced by multiple quantitative trait loci, their interaction, the environment, and the interaction between QTL and environment. Linkage analysis and association mapping are the two most commonly used tools for dissecting genetic basis of complex traits. Linkage analysis in plants typically localizes QTLs to 10 to 20 cM intervals because of the limited number of recombination events that occur during the construction of mapping populations and the cost for propagating and evaluating a large number of lines (Holland, 2007; Zhu *et al.*, 2008). Association mapping resolve complex trait variation down to the sequence level by exploiting historical and evolutionary recombination events at the population level (Nordborg and Tavare, 2002). Association mapping offers three advantages over traditional linkage analysis, (a). higher mapping resolution, (b). greater allele number and (c). reduced research time. Since its introduction to plants, it has continued to gain popularity and preference in genetic research because of increasing interests in scientific community to identify novel and superior alleles, recent advances in high throughput genomic technologies, and improvements in statistical methods.

It falls into two broad categories. First, candidate gene association mapping, which is based on the polymorphisms in, selected candidate genes responsible for phenotypic variation for specific traits. Second, genome-wide association mapping or also referred to as genome which searches for genetic variation in the whole genome to find QTLs associated with various complex traits. Until now, a number of experiments focusing on LD and association mapping have been published in number of crop plants. The association mapping studies in important crop plants has been listed in Table 1.

### Practical approaches to association mapping

The various steps in the identification and characterization of quantitative trait loci from the natural population through association mapping for use in marker assisted selection are presented in Figure 1.

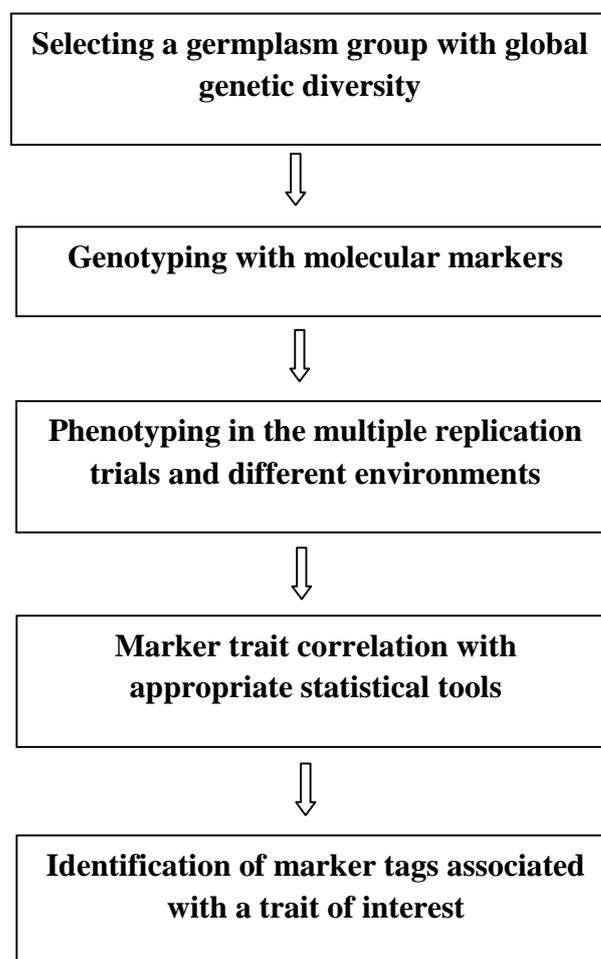


Figure 1. The various steps involved association mapping

Table 1. List of important association mapping studies in crop plants

Crop plant	Population	Sample size	Traits	Reference
Rice	Diverse accessions	103	Yield and component traits	Agrama <i>et al.</i> , 2007
	Diverse landraces	577	Starch quality	Bao <i>et al.</i> , 2006
Wheat	Diverse cultivars	95	Kernel size, milling quality	Breseghello and Sorrells, 2006
	Germplasm accessions	108	Morphological and yield traits	Yao <i>et al.</i> , 2009
Maize	Elite inbred lines	71	Flowering time	Andersen <i>et al.</i> , 2005
	Diverse inbred lines	95	Flowering time	Salvi, 2007
	Diverse inbred lines	102	Kernel composition	Wilson <i>et al.</i> , 2004
	Elite inbred lines	75	Kernel color	Palaisa <i>et al.</i> , 2004
	Diverse inbred lines	57	Sweet taste	Tracy <i>et al.</i> , 2006
	Elite inbred lines	553	Oleic acid content	Belo <i>et al.</i> , 2008
Sorghum	Diverse inbred lines	282	Carotenoid content	Harjes <i>et al.</i> , 2008
	Diverse inbred lines	377	Community resource report	Casa <i>et al.</i> , 2008
Barley	Diverse cultivars	148	Days to heading, leaf rust, yellow dwarf virus	Kraakman <i>et al.</i> , 2006
Potato	Diverse cultivars	123	Late blight resistance	Malosetti <i>et al.</i> , 2007
	Diverse cultivars	221	Yield and quality traits	D'hoop <i>et al.</i> , 2008
Sugarcane	Diverse clones	154	Disease resistance	Wei <i>et al.</i> , 2006
Pearl millet	Inbred lines / accessions	90/598	Morphological and yield traits	Saidou <i>et al.</i> , 2009
Soybean	Breeding lines	139/115	Yield and quality traits	Wang <i>et al.</i> , 2008
	Germplasm accessions	96	Yield and quality traits	Jun <i>et al.</i> , 2008
Sugar beet	Inbred lines	111	Morphological traits	Stich <i>et al.</i> , 2008
	Elite clones	768	Morphological and yield traits	Stich <i>et al.</i> , 2008

## Genotyping

Association mapping studies commonly use unlinked and putatively neutral markers to characterize genetic variation in the accessions used in the mapping study and to account for population structure. There are many types of markers that can be used for this, which includes AFLP, RAPD, SSRs and SNPs. The dominant marker like AFLP cannot be distinguishing between heterozygous and homozygous genotypes. This poses problems when using AFLP markers for estimating population structure or for use directly in mapping. On the other hand, SSR markers are codominant and highly polymorphic. The high variability of SSR markers, coupled with the availability of semi-automatic detection methods, have made them the markers of choice for use in estimating population structure or pair wise relatedness among individuals.

## Candidate Genes

Candidate-gene association mapping is a hypothesis driven approach to dissect complex traits, with biologically relevant candidates selected and ranked based on the evaluation of available results from genetic, biochemical, or physiology studies in crop plants. This method requires the identification of SNPs between lines and within specific genes. The most common method of identifying candidate gene SNPs relies on the resequencing of amplicons from several genetically distinct individuals of a larger association population. In the SNP discovery panel, fewer diverse individuals are needed to identify common SNPs and large number of diverse individuals is needed to identify rarer SNPs (Zhu *et al.*, 2008).

## Whole-Genome Scan

For whole-genome association scans to be performed in crops, the first step is to make use of high-capacity DNA sequencing instruments or high-density oligonucleotide (oligo) arrays for efficient identification of SNPs at a density that exactly reflects genome-wide LD structure and haplotype diversity. Large numbers of SNP markers are needed for powerful whole-genome scans in plant species with low LD and high haplotype diversity. After SNPs are identified, different array-based platforms can be used to genotype thousands of tag SNPs in parallel (Zhu *et al.*, 2008).

## Candidate genes versus whole genome scans

The extent of LD determines mapping resolution and the numbers of markers needed for a sufficient coverage of the genome. Hence, extent of LD in the organism of interest is the most important factor when deciding between a candidate gene and a whole genome approach. One should also take into account the variation of recombination rates across the genome when considering the extent of LD. Relatively few markers are needed in species where LD extends over long physical distances for adequate genome coverage (Hall *et al.*, 2010). For example, in *Arabidopsis thaliana*, LD can extend for tens or even hundreds of kilo base pairs. Thus, genome scan can be done with less number of evenly spaced SNPs markers. Whereas, in many predominantly or obligately cross pollinating species like maize and many forest trees, LD extends only a few hundred base pairs at the most and thus,

genome scan would require millions of SNPs. Since association mapping in candidate gene approach is restricted to relevant candidates genes assumed to be involved in controlling the expression of trait of interest that is why it is hypothesis-driven than a genome scan. The selection of candidates is not straightforward when it is based on the information obtained from genetic, physiological or biochemical studies in crop plants. But, it is straightforward when restricted to well characterized developmental pathways or to traits with a well-understood biochemical basis. Candidate gene studies are less demanding in terms of the number of markers that are required and many candidate gene association studies have successfully been completed using tens to hundreds of markers in mapping populations consisting of a few hundred individuals. However, it is important to remember that a candidate gene approach is limited by the choice of candidate genes that are identified and hence always runs the risk of missing out on identifying causal mutations that are located in non-identified candidate genes (Hall *et al.*, 2010).

## Phenotyping

The Phenotyping has received attention the importance of accurate identification and scoring of genotypes is being affected by Phenotyping. It has been shown, that increasing the number of individuals phenotyped is far more efficient than increasing the number of SNPs for increasing the power in association studies. A typical association mapping study usually involves a diverse set of accessions and phenotypic of individual accessions within a site in replication is usually needed to increase precision in phenotypic measurements, by eliminating environmentally induced noise and measurement errors.

## Population structure control

The risk of incurring false positives due to population structure is one of the main hurdles for using association mapping to dissect the genetic architecture of complex traits in crop plants. When the phenotypic trait that is also correlated with the underlying population structure at neutral loci will show an inflated number of positive associations results in the problem of population structure. The problem of population structure is well known and many methods have been developed to deal with this problem. Some important statistical software packages used in association mapping studies have been listed in Table 2.

## Nested Association Mapping (NAM)

Linkage analysis identifies broad chromosome regions of interest with relatively low marker coverage, whereas, association mapping provide high resolution mapping with either prior information on candidate genes or a genome scan with very high marker coverage. It is desirable to combine the advantages of both linkage analysis and association mapping to develop alternate integrated mapping strategy ultimately to improve the mapping resolution without requiring excessively dense marker maps. So both linkage and LD mapping have been proposed as a fine mapping approach. Nested Association Mapping (NAM) is practically used in maize crop (*Zea mays* L.), as a more powerful strategy for dissecting the genetic basis of complex traits in species with low LD (Yu *et al.*, 2008).

Table 2. Important statistical software packages used in association mapping studies

Software package	Website	Reference
STRUCTURE	<a href="http://pritch.bsd.uchicago.edu/structure.html">http://pritch.bsd.uchicago.edu/structure.html</a>	Pritchard <i>et al.</i> 2000
EIGENSOFT	<a href="http://genepath.med.harvard.edu/reich/Software.htm">http://genepath.med.harvard.edu/reich/Software.htm</a>	Price <i>et al.</i> 2006
EMMA	<a href="http://mouse.cs.ucla.edu/emma/">http://mouse.cs.ucla.edu/emma/</a>	Kang <i>et al.</i> 2008
STRAT	<a href="http://pritch.bsd.uchicago.edu/software/STRAT.html">http://pritch.bsd.uchicago.edu/software/STRAT.html</a>	Pritchard <i>et al.</i> 2000
TASSEL	<a href="http://www.maizegenetics.net">http://www.maizegenetics.net</a>	Bradbury <i>et al.</i> 2007
SAS	<a href="http://www.sas.com">http://www.sas.com</a>	SAS Institute, 1999
MTDFREML	<a href="http://aipl.arsusda.gov/curtvt/mtdfreml.html">http://aipl.arsusda.gov/curtvt/mtdfreml.html</a>	Boldman <i>et al.</i> , 1993
ASREML	<a href="http://www.vsni.co.uk/products/asreml">http://www.vsni.co.uk/products/asreml</a>	Gilmour <i>et al.</i> , 2002

Advantages of NAM are: (a). it is less sensitivity to genetic heterogeneity; (b). higher resolution power; and (c). Higher efficiency in using the genome sequence or dense markers while still maintaining high allele richness due to diverse founders. This is in striking contrast to *Arabidopsis* and rice, where large-effect QTLs have been found in many studies.

### Conclusion

World is the still the reservoir of the wild and unadapted germplasm that may hold key to increase productivity in future in crop plants. The progenitor species constitute the largest source of still unfolded variability for complex traits. The novel QTLs identified are good candidates for fine mapping and positional cloning studies, while, the QTLs that are mapped to regions consistent with other studies can be useful for marker assisted transfer of these QTLs. These QTL mapping studies suggest that the genetic basis of complex traits is influenced by breeding system, population size, population structure and selective history. Natural populations possess a stunning diversity of phenotypic variation for yield and yield related traits. This phenotypic variation is due to multiple interacting loci which are sensitive to the environmental conditions. Knowledge of the relationship between DNA sequence variation and variation in phenotypes for these quantitative or complex traits will definitely help in future for increasing the speed of selective breeding programmes in agriculturally important crop plants and for predicting adaptive evolution in crops.

### REFERENCES

- Agrama, H.A., Eizenga, G.C. and Yan W. 2007. Association mapping of yield and its components in rice cultivars. *Mol. Breed.* 19(4):341–356.
- Andersen, J.R., Schrag, T., Melchinger, A.E., Zein, I. and Lubberstedt, T. 2005. Validation of Dwarf polymorphisms associated with flowering time in elite European inbred lines of maize (*Zea mays* L.). *Theor. Appl. Genet.* 111(2):206–217.
- Bao, J.S., Corke, H. and Sun, M. 2006. Microsatellites, single nucleotide polymorphisms and a sequence tagged site in starch-synthesizing genes in relation to starch physicochemical properties in nonwaxy rice (*Oryza sativa* L.). *Theor. Appl. Genet.* 113(7):1185–1196.
- Belo, A., Zheng, P., Luck, S., Shen, B., Meyer, D.J., Li, B., Tingey, S. and Rafalski, A. 2008. Whole genome scan detects an allelic variant of *fad2* associated with increased oleic acid levels in maize. *Mol. Genet. Genomics* 279:1–10.
- 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.
- Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y. and Buckler, E.S. 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23:2633–2635.
- Breseghele, F. and Sorrells, M.E. 2006. Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics* 172:1165–1177.
- Casa, A.M., Pressoira, G., Brown, P.J., Mitchell, S.E., Rooney, W.L., Tuinstra, M.R., Franks, C.D. and Kresovich, S. 2008. Community resources and strategies for association mapping in sorghum. *Crop Sci.*, 48:30–40.
- Dhoop, B.B., Paulo, M.J., Mank, R.A., van Eck, H.J. and van Eeuwijk, F.A. 2008. Association mapping of quality traits in potato (*Solanum tuberosum* L.). *Euphytica* 161:47–60.
- Gilmour, A.R., B.J. Gogel, B.R. Cullis, S.J. Welham, and R. Thompson. 2002. ASREML user guide release 1.0. VSN International Ltd., Hemel Hempstead, UK.
- Gupta, P.K. Rastgi, S. and Kilwal, P.L. 2005. Linkage disequilibrium and association studies in higher plants: present status and future prospects. *Plant Mol. Biol.*, 57: 461–485.
- Hall, D., Tegstrom, C. and Ingvarsson, P.K. 2010. Using association mapping to dissect the genetic basis of complex traits in plants, *Brief. Funct. Genomics*. 9(2):157–165.
- Harjes, C.E., Rocheford, T.R., Bai, L., Brutnell, T.P., Kadianis, C.B., Sowinski, S.G., Stapleton, A.E., Vallabhaneni, R., Williams, M., Wurtzel, E.T., Yan, J.B. and Buckler, E.S. 2008. Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. *Science*, 319:330–333.
- Holeyachi, P., Purohit, J., Devi E.L. and Gururaj N. 2013. Deciphering genetic basis of complex traits for crop improvement. *Res. J. Agriculture and Forestry Sci.*, 1(9):14–18.
- Holland, J.B. 2007. Genetic architecture of complex traits in plants. *Curr. Opin. Plant Biol.*, 10: 156–161.
- Jun, T.H., Van, K., Kim, M.Y., Lee, H.S. and Walker, D.R. 2008. Association analysis using SSR markers to find QTL for seed protein content in soybean. *Euphytica* 162:179–191.

- Kang, H.M., Zaitlen, N.A., Wade, C.M., 2008. Efficient control of population structure in model organism association mapping. *Genetics* 178:1709–23.
- Kraakman, A.T.W., Martinez, F., Mussiraliev, B., Eeuwijk, F.A. and Niks, R.E. 2006. Linkage disequilibrium mapping of morphological, resistance, and other agronomically relevant traits in modern spring barley cultivars. *Mol. Breed.* 17:41–58.
- Malosetti, M., vander Linden, C.G., Vosman, B. van Eeuwijk, F.A. 2007. A mixed-model approach to association mapping using pedigree information with an illustration of resistance to *Phytophthora infestans* in potato. *Genetics* 175:879–889.
- Mauricio R. 2001. Mapping quantitative trait loci in plants: uses and caveats for evolutionary biology, *Nat Rev Genet.*, 2:370–381.
- Nordborg, M. and Tavaré, S. 2002. Linkage disequilibrium: what history has to tell us. *Trends Genet.* 18:83–90.
- Palaisa, K., Morgante, M., Tingey, S. and Rafalski, A. 2004. Long-range patterns of diversity and linkage disequilibrium surrounding the maize Y1 gene are indicative of an asymmetric selective sweep. *Proc. Natl. Acad. Sci. USA* 101:9885–9890.
- Price A.L., Patterson, N.J., Plenge R.M. et al. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.*, 38:904–9.
- Pritchard, J.K., Stephens, M. and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155:945–959.
- Saidou, A.A., Mariac, C., Luong, V., Pham, J.L., Bezancon, G. and Vigouroux, Y. 2009. Association studies identify natural variation at PHYC linked to flowering time and morphological variation in pearl millet. *Genetics* 182:899–910.
- Salvi, S. 2007. Conserved non-coding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proc. Natl. Acad. Sci. USA* 104:11376–11381.
- SAS Institute. 1999. SAS/STAT user's guide. Version 8 SAS Institute, Inc, Cary, NC.
- Stich, B., Melchinger, A.E., Mbhring, M.J., Schechert, A. and Piepho, H.P. 2008a. Association mapping in multiple segregating populations of sugar beet (*Beta vulgaris* L.). *Theor. Appl. Genet.* 117(7):1167–1179.
- Stich, B., Piepho, H.P., Schulz, B. and Melchinger, A.E. 2008b. Multi-trait association mapping in sugar beet (*Beta vulgaris* L.). *Theor Appl Genet* 117(6):947–954.
- Tracy, W.F., Whitt, S.R. and Buckler, E.S. 2006. Recurrent mutation and genome evolution: Example of Sugary1 and the origin of sweet maize. *Crop Sci.* 46:S49–S54.
- Veerasha, B. A., Rudra Naik, V., Chetti, M. B., Desai, S. A. and Suma S. Biradar 2015. QTL mapping in crop plants: principles and applications *Int. J. Development Res.* 5(1):2961-2965.
- Wang, J., Mc Clean, P.E., Lee, R., Goos, J. and Helms, T. 2008. Association mapping of iron deficiency chlorosis loci in soybean (*Glycine max* L. Merr.) advanced breeding lines. *Theor Appl Genet* 116:777–787.
- Wei, X.M., Jackson, P.A., McIntyre, C.L., Aitken, K.S. and Croft, B. 2006. Associations between DNA markers and resistance to diseases in sugarcane and effects of population substructure. *Theor. Appl. Genet.* 114:155–164.
- Wilson, L.M., Whitt, S.R., Rocheford, T.R., Goodman, M.M. and Buckler, E.S. 2004. Dissection of maize kernel composition and starch production by candidate gene association. *Plant Cell* 16:2719–2733.
- Yao, J., Wang, L., Liu, L., Zhao, C. and Zheng, Y. 2009. Association mapping of agronomic traits on chromosome 2A of wheat. *Genetica* 137:67–75.
- Yu, J., Holland, J.B., Mc Mullen, M.D. and Buckler E.S. 2008. Genetic design and statistical power of nested association mapping in maize. *Genetics* 178:539–551.
- Zhu, C., Gore, M., Buckler, E.S. and Yu, J., 2008. Status and prospects of association mapping in plants, *The Plant Genome.*, 1(1): 5-20.

\*\*\*\*\*