Research Journal of Agriculture and Forestry Sciences _____ ISSN 2320-6063
Vol. **1(9),** 14-18, October **(2013)** *Res. J. Agriculture and Forestry Sci.*

Review Paper

Deciphering Genetic Basis of Complex Traits for Crop Improvement

Praveen Holeyachi[1]*, Jyotika Purohit[2], Lamalakshmi Devi E.[1] and Gururaj N.[3]
[1]Department of Genetics and Plant Breeding, GBPUAT, Pantnagar - 263 145, Uttarakhand, INDIA
[2]Department of Plant Pathology, GBPUAT, Pantnagar - 263 145, Uttarakhand, INDIA
[3]National Fertilizer Limited, INDIA

Available online at: www.isca.in, www.isca.me
Received 14th September 2013, revised 22nd September 2013, accepted 6th October 2013

Abstract

In current biology, a greater challenge is to know the genetic basis of complex traits or quantitative traits which is controlled by the cumulative effects of quantitative trait loci (QTLs), epistasis, environment and interaction between quantitative trait locus (QTL) and environment. Availability of molecular markers and linkage maps have made it possible to understand the genetic basis of complex traits through the marker-based mapping to locate the chromosomal regions or QTLs of interest. Linkage analysis based mapping, association mapping and nested association mapping (NAM) are most commonly used methods for understanding the genetic basis underlying quantitative variation which will help in improving traits such as yield, nutritional quality and resistance to abiotic and biotic stress by developing new insights and methodologies.

Keywords: genetic, complex traits, quantitative trait loci, yield, nutritional quality, resistance, abiotic and biotic stress

Introduction

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 traits of interest in plant breeding (*e.g.*, yield, height, drought resistance, disease resistance in many species, etc.) are quantitative or polygenic or continuous or complex traits. The multiple-factor hypothesis was proposed by Nillson-Ehle in 1909. Since then 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). Several QTLs regulate the expression of a single trait. QTL refers to a region of DNA linked with a particular trait and QTLs refers to the situation when two or more regions of DNA from the same or different chromosomes are linked with a particular trait.

Polygenic traits do not follow patterns of Mendelian inheritance unlike oligoenic traits. Instead, they show continuous variation depicted by a bell curve. These complex traits complicate the works of breeders because performance only partially reflects the genetic values of the individuals[1]. Genetic basis refers to the numbers and genome locations of genes that affect a trait, the magnitude of their effects and the relative contributions of additive, dominant and epistatic gene effects. In current biology, a greater challenge is to know the genetic basis of complex traits or quantitative traits.

The principles of mapping quantitative trait loci (QTLs) that affect the natural variation in complex traits by linkage to polymorphic marker loci with Mendelian segregation have been known since the early twentieth century. The genetic dissection of complex traits was limited to a few model organisms until the late 1980s due to dearth of polymorphic markers. Since then, the discovery of abundant molecular markers, advances in rapid and cost-effective genotyping methods and the development of statistical methods for QTL mapping have revolutionized the field of mapping quantitative traits. In 1989, the landmark paper by Lander and Botstein stimulated a number of studies on QTL mapping. The assumption in QTL mapping is that QTLs can be localized to visible marker loci through their genetic linkage with genotypes which we can easily classify. If a QTL is linked to a marker locus, then individuals with different marker locus genotypes will have different mean values of the quantitative trait[2]. The two general goals of QTL mapping in plants are to (a) improve our knowledge regarding inheritance and genetic architecture of quantitative traits, and (b) search markers that can be used for indirect selection in plant breeding[3].

For QTL mapping, there are various requirements. First, mapping populations which may be experimental populations for linkage-based mapping or natural/breeding populations for association mapping are required in which phenotyping for the trait(s) of interest is done. Another requirement is molecular marker(s) for genotyping. Thus, using these molecular markers linked to the trait(s) of interest are identified using statistical programs. Finally, test is done for the applicability and reliability of the markers associated with major QTLs in predicting the trait(s) in related families (marker validation or verification) for QTLs of medium to large effect[1].

QTL Mapping

Once the phenotype and genotype data are generated, scientists are curious to test the two hypotheses in QTL mapping.

According to null hypothesis (H0), no QTL is present or a QTL is present but it is not linked to the marker(s) and according to alternative hypothesis (HA), a QTL is present and it is linked to the marker(s). Different statistical methods are available for testing the two hypotheses. These methods can be classified into three groups based on the type of population(s) for mapping: i. methods which require the development of appropriate mapping population(s) using experimentally designed crosses, for example, analysis of variance, simple interval mapping (SIM), composite interval mapping (CIM), multiple interval mapping (MIM); ii. methods that require natural or breeding populations, for example, linkage disequilibrium based mapping and iii. methods that use either appropriate mapping populations developed by designed crosses or natural or breeding populations, for example, principal component analysis based mapping and partial least square regression. The statistical methods employed for QTL analysis can also be categorized into two groups based on their requirements for genetic maps: i. methods that does not require prior construction of genetic linkage map, for example, analysis of variance, linkage disequilibrium-based mapping, principal component analysis-based mapping and partial least squares regression; ii. methods that require availability of genetic map for the population, for example, simple interval mapping, composite interval mapping and multiple interval mapping. For the latter, prior to QTL analysis, researchers need to construct a genetic linkage map for the population by conducting linkage analyses on the genotypic data[4].

Linkage analysis-based QTL mapping

Interval mapping (IM) which is often called simple interval mapping (SIM) is a more powerful QTL mapping method developed by Lander and Botstein in 1989. As an improvement over single-QTL models, Multiple-QTL models were developed which have the ability to separate linked QTLs on the same chromosome and also detect interacting QTLs that may otherwise go undetected. Various approaches have been proposed for mapping multiple QTLs. Composite interval mapping (CIM) was proposed combining SIM with multiple regression analysis in mapping[5,6,7]. For mapping multiple QTLs simultaneously, multiple interval mapping (MIM) was proposed and implemented[8]. The basic idea of multiple interval mapping (MIM) is to search, test and estimation of positions, effects and interactions of multiple QTLs by fitting multiple putative QTL effects and epistatic effects directly in a model. MIM tends to be more powerful and precise in detecting QTLs as compared to SIM and CIM.

QTL mapping is extremely useful in detecting various regions of genome that affect the expression of complex traits in a large number of species[9]. However, it suffers from number of limitations. First, as only two parents are used to construct QTL mapping populations, allelic variation in each cross is restricted. Second, recombination events per chromosome are usually less since early generation crosses are used, thus, limiting the resolution of the genetic map. A typical QTL identified from a cross consisting of a few hundred offspring can span anywhere between a few to ten cMs corresponding to genomic regions comprising several megabases. The process of identifying the causal gene in a QTL region is a tedious and quite time consuming task in such large genome regions as they contain hundreds of genes[10]. Apart from this, the development of mapping populations is either not possible or is very time consuming in many organisms. For example, the long generation time of most forest trees have resulted in prevention or slowing down of any progress in dissecting the genetic basis of complex traits using QTL mapping[11]. Association or linkage disequilibrium mapping have been hailed as a more efficient way of determining genetic basis of complex traits.

Association Mapping

Association mapping, also known as linkage disequilibrium (LD) mapping, has become a tool in recent past to study complex trait variation by making use of historical and evolutionary recombination events in the population. Association mapping offers three advantages over traditional linkage analysis, i. higher mapping resolution, ii. reduced research time, and iii. greater allele number[12]. Since its introduction to plants[13], 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 many number of plant species. Many major crops, such as maize (*Zea mays,* L.), soybean (*Glycine max* (L.) Merr.), tomato (*Lycopersicon esculentum* Mill), barley (*Hordeum vulgare* L.), sorghum (*Sorghum bicolor* (L.) Moench), wheat (*Triticum aestivum* L.), and potato (Solanum *tuberosum* L.). In this regard, substantial work has also been done in tree species such as aspen (*Populus tremula* L.) and loblolly pine (*Pinus taeda* L.)[14].

Initiation of Association Mapping

Different factors like genetic aspects of the species and the associated germplasm should be carefully consider by scientists before initiating association mapping. If the population consists wild accessions obtained from a germplasm bank, then ploidy level of individuals should be evaluated. This helps in avoiding the difficulty faced in differentiating the effects of functional polymorphisms from that of allele dosage. It will be useful if we examine various genetic tools available for a given species

because developing and studying an association mapping population requires a long-term commitment[14]. Genetic diversity, extent of genome-wide LD, and relatedness within the population determine the mapping resolution, marker density, statistical methods, and mapping power.

Candidate Genes

Candidate-gene association mapping is a hypothesis driven approach to complex trait dissection, with biologically relevant candidates selected and ranked based on the evaluation of available results from genetic, biochemical, or physiology studies in model and non-model plant species[15]. 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 are needed to identify rarer SNPs[14].

Whole-Genome Scan

For whole-genome association scans to be performed in crops, 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 number of SNP markers are needed for powerful whole-genome scans in plant species with low LD and high haplotype diversity, for example, maize and sunflower. After SNPs are identified, different array-based platforms can be used to genotype thousands of tag SNPs in parallel[14].

Candidate genes versus whole genome scans

Extent of LD in the organism of interest is the most important factor when deciding between a candidate gene and a whole-genome approach, because the extent of LD determines mapping resolution and the numbers of markers needed for a sufficient coverage of the genome[16]. 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[17]. 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[11], 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 which 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 both model and non-model plant species[11]. 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[17].

Nested Association Mapping (NAM)

Two commonly used approaches to study the genetic basis of complex traits are linkage based QTL mapping and association mapping[15]. 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 would be wise if we 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[18]. Such a possibility exists for the maize crop (*Zea mays* L.), because of the availability of a highly diverse collection of germplasm and the feasibility of creating segregating progenies and immortal genotypes through self-fertilization.

Family based pedigrees are valuable in maize (*Zea mays*), which has high levels of outcrossing and a large effective population size. This results in very low linkage disequilibrium, which decays within hundreds of nucleotides in most populations. Using present technology available at hand, it is very expensive to record polymorphisms at this density which makes genome scan, a challenging task in maize[19]. A different type of family breeding design known as the Nested Association Mapping (NAM) population has been used in maize which is an outcome of a large collaboration among maize geneticists[20]. In NAM strategy, a common mapping resource is generated which enables researchers to efficiently utilize genetic, genomic, and systems biology tools. Advantages of NAM are: i. it is less sensitivity to genetic heterogeneity; ii. higher resolution power; and iii. higher efficiency in using the genome sequence or dense markers while still maintaining high allele richness due to diverse founders.

A recent study examining flowering time in nearly 1 million plants from around 5,000 NAM recombinant inbred lines found that the genetic architecture of flowering time was highly polygenic[20]. Around 50 loci appeared to contribute to variation in flowering time, with many loci showing small, nearly additive effects. This is in striking contrast to *Arabidopsis* and rice, where large-effect QTLs have been found in many studies.

Multiparent Advanced Generation Inter-Cross (MAGIC) Lines

The association mapping, nested association mapping and multiparent intercross populations which are second generation mapping resources, potentially address the major drawbacks of available mapping resources. Both linkage and association analysis can be conducted in such populations without encountering the limitations of structured populations. For developing MAGIC lines, use of eight founders and a fixed population of 1000 individuals are most appropriate[21]. For example in *Arabidopsis*, to develop MAGIC population, 19 founding genotypes were crossed together for four generations to increase the level of recombination, followed by six generations of self-pollination to develop 342 quasi-independent recombinant inbred lines. This population was then used study flowering time and other complex traits. As a result, number of QTLs were identified near known candidate genes[22], including the flowering time genes *FRIGIDA* and *FLOWERING LOCUS C*, which also were evident in the genome wide scan[23]. In comparison to population-based mapping, pedigree approaches like MAGIC lines can avoid complications of historical population structure, although QTLs cannot be resolved to regions of a few genes.

Limitations with MAGIC populations are that they require longer time and more resource to be generated. They show extensive segregation for developmental traits, limiting their use in the analysis of complex traits. However, multiparent intercross population resources are likely to bring a paradigm shift towards QTL analysis in plant species[21].

Conclusion

These QTL mapping studies suggest that the genetic basis of complex traits is influenced by breeding system, population size, selective history, and population demography. 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[24]. 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 plants and for predicting adaptive evolution[25].

References

1. Semagn K., Bjørnstad A. and Xu Y., The genetic dissection of quantitative traits in crops, *Electron. J. Biotechnol*., **13(5)**, 14 **(2010)**

2. Lander E.S. and Botstein D., Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps, *Genetics,* **121(1)**, 185-199 **(1989)**

3. Bernardo R., Molecular markers and selection for complex traits in plants: learning from the last 20 years, *Crop Science.,* **48(5)**, 1649-1664 **(2008)**

4. Semagn K., Bjørnstad A. and Ndjiondjop M.N., Principles, requirements and prospects of genetic mapping in plants, *African Journal of Biotechnology*, **5(25)**, 2569-2587 **(2006)**

5. Jansen R.C., Interval mapping of multiple quantitative trait loci, *Genetics.,* **135(1)**, 205-211 **(1993)**

6. Zeng Z., Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci, *Proceedings of the National Academy of Sciences*, **90(23)**, 10972-10976 **(1993)**

7. Zeng Z., Precision mapping of quantitative trait loci, *Genetics,* **136(4)**, 1457-1468 **(1994)**

8. Kao C.H., Zeng Z. and Teasdale R.D., Multiple interval mapping for quantitative trait loci, *Genetics,* **152(3)**, 1203-1216 **(1999)**

9. Mauricio R., Mapping quantitative trait loci in plants: uses and caveats for evolutionary biology, *Nat Rev Genet*., **2**, 370–381 **(2001)**

10. Price A.H., Believe it or not, QTLs are accurate, *Trends Plant Sci*., **11,** 213–216 **(2006)**

11. Neale D.B. and Savolainen O., Association genetics of complex traits in conifers, *Trends Plant Sci*., **9**, 325–330 **(2004)**

12. Yu J. and Buckler E.S., Genetic association mapping and genome organization of maize, *Current Opinion in Biotechnology*, **17(2)**, 155-160 **(2006)**

13. Thornsberry J.M., Goodman M.M., Doebley J., Kresovich S., Nielsen D. and Buckler E.S., Dwarf8 polymorphisms associate with variation in flowering time, *Nat. Genet*., **28,** 286–289 **(2006)**

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

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

16. Whitt S.R. and Buckler E.B., Using natural allelic diversity to evaluate gene function, *Methods Mol Biol*., **236**, 123–40 **(2003)**

17. Hall D., Tegstrom C. and Ingvarsson P.K., Using association mapping to dissect the genetic basis of complex traits in plants, *Brief. Funct. Genomics*., **9(2)**, 157-165 **(2010)**

18. Yu J., Holland J.B., McMullen M.D. and Buckler E.S., Genetic Design and Statistical Power of Nested Association Mapping in Maize, *Genetics,* **178,** 539–551 **(2008)**

19. Mitchell-Olds T., Complex-trait analysis in plants, *Genome Biology,* **11(4)**, 113 **(2010)**

20. Buckler E.S., Holland J.B., Bradbury P.J., Acharya C.B., Brown P.J., Browne C., Ersoz E., Garcia S.F., Garcia A., Glaubitz J.C., Goodman M.M., Harjes C., Guill K., Kroon D.E., Larsson S., Lepak N.K., Li H., Mitchell S. E., Pressoir G., Peiffer J.A., Rosas M.O., Rocheford T.R., Romay M.C., Romero S., Salvo S., Villeda H.S., Silva H.S., Sun Q., Tian F., Upadyayula N., Ware D., Yates H., Yu J., Zhang Z., Kresovich S. and McMullen M.D., The genetic architecture of maize flowering time, *Science.,* **325(5941)**, 714-718 **(2009)**

21. Sujay R., Arunita R. and Patil J.V., Multiparent intercross populations in analysis of quantitative traits, *J. Genet.*, **91(1)**, **(2012)**

22. Kover P.X., Valder W., Trakalo J., Scarcelli N., Ehrenreich I.M., Purugganan M.D. *et al.,* A multiparent advanced generation inter–cross to fine map quantitative traits in *Arabidopsis thaliana*, *PLoS Genet.,* **5(7)**, e1000551 **(2009)**

23. Atwell S., Huang Y.S., Vilhjálmsson B.J., Willems G., Horton M., Li Y. *et al.,* Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines, *Nature,* **465**, 627–631 **(2010)**

24. Falconer D.S. and Mackay T.F.C., Introduction to Quantitative Genetics (Addison Wesley Longman, Harlow), **(1996)**

25. Mackay T.F.C., Stone E.A. and Ayroles J.F., The genetics of quantitative traits: challenges and prospects, *Nat. Rev. Genet.,* **10**, 565-577 **(2009)**