ICAR Short Course on

Molecular Approaches for Cotton Improvement

September 19-28, 2019

COMPENDIUM OF LECTURES and TRAINING MANUAL

Course Director : Dr. V. N. Waghmare

Course Coordinators: Dr. Raghavendra K.P. and Dr. Santosh, H. B

ICAR-CENTRAL INSTITUTE FOR COTTON RESEARCH

Post Bag No. 2, Shankarnagar PO, Nagpur 440 010

Molecular Markers for Crop Improvement

Santosh, H. B.[1]*, A. Manivannan[2] and Raghavendra, K. P.[1]
[1]ICAR-Central Institute for Cotton Research (CICR), Nagpur
[2]ICAR- Central Institute for Cotton Research, RS, Coimbatore
*Email ID: santoshgpb@gmail.com

Molecular Marker Systems

Genetic information in the form of DNA is stored in cell at three places, *viz.*, nucleus, chloroplast and mitochondria. The genetic variation arises in this stored information due to the change in the DNA sequence by base substitution, deletion, insertion, duplication and inversion. These variations are manifested as variation in trait expression (at phenotype level) or amino acid sequence variation (at protein level) or nucleotide sequence variation (at DNA level). In general, marker is anything that marks or tags or identify an individual. Any heritable variation following Mendelian inheritance can be called as genetic marker. Genetic markers facilitate mapping and discovery of valuable genes hence, they are widely used in crop improvement programmes. Compared to conventional phenotypic selection, marker assisted selection (MAS) provides more reliability, efficiency and rapidity in plant breeding programmes. Genetic markers are of three types, *viz.*, morphological markers, biochemical markers and DNA based markers.

I. **Morphological markers**: These are related to shape, size, colour and surface of various plant parts.
II. **Biochemical marker**: Such markers are related to variation in protein and amino acid banding patterns.
III. **DNA markers**: Typically, small regions of DNA showing sequence polymorphism in different individuals with in a species, e.g. Restriction Fragment Length Polymorphism- RFLPs; Amplified Fragment Length Polymorphism- AFLP; Random Amplified Polymorphic DNA- RAPD; Cleaved Amplified Polymorphic Sequence-CAPS; Simple Sequence Repeats- SSRs; Single Nucleotide Polymorphism (SNP), etc.

Characteristics of ideal genetic marker:

a. High level of polymorphism
b. Stability, no influence of environment
c. Simplicity of observation
d. Non-epistatic
e. Co-dominancy
f. Genome-wide distribution
g. Low cost
h. Reproducibility
i. Portability between species

j. Mendelian inheritance

Morphological and biochemical markers are have limited applications in crop improvement compared to DNA based markers because they are less in number, not well distributed across the genome, stage or/and tissue specific expression, influenced by external environment, not amenable to complete automation and limited polymorphism. Hence, widely used DNA markers are detailed in length in this lecture notes. There are different classes of DNA markers. Based on their evolution, first generation markers (hybridization based markers like RFLPs); second generation markers (most of PCR based markers like RAPD, AFLP, SSRs) and third generation markers (sequence based markers like SNPs).

Types of DNA markers

i) Restriction Fragment Length Polymorphism (RFLP):

RFLP is hybridization based marker, developed by Botstein *et al.*, (1980). This is the polymorphism which is detected based on the differential hybridization of cloned DNA to DNA fragments in a sample of restriction enzyme digested DNAs. Hence, the polymorphism detected in the fragment length is due to variation in restriction site(s) of the enzyme used and these RFLPs are defined by specific enzyme-probe combinations.

Strength:

- ✓ Robust and reproducible
- ✓ Co-dominant nature and hence can be effectively used to differentiate homozygotes from the heterozygotes
- ✓ Particularly useful in comparative genome mapping

Weakness:

- The assay is tedious and time consuming
- Requires large quantities of high quality DNA
- Not amenable to complete automation
- Associated with health hazards as it used radioactivity

ii) Random Amplified Polymorphic DNA (RAPD):

RAPD markers were given by Williams *et al.*, (1990) and are based on the differential PCR amplification of a sample of DNAs from random oligo-nucleotide sequence. RAPDs are genetically dominant in nature. RAPD uses single random primer usually of 10 nucleotides long, to amplify inverted repeats in the genome and can anneal at multitude of genomic locations.

Strength:

- ✓ Fast and so called Rapid markers
- ✓ Requires less quantities of DNA
- ✓ Needs limited investment in time and training
- ✓ Commercially available

- ✓ No radioactivity involved
- ✓ No need of sequence information

Weakness:

- Lack of reproducibility in markers patterns across lab and experiments
- Dominant in nature

iii) Amplified Fragment Length Polymorphism (AFLP):

AFLP markers were given by Vos *et al.*, (1995) and these polymorphism are generated using a procedure that combines restriction digestion and PCR amplification. The basic procedure is as follows:

- ➢ Digestion of genomic DNA with a combination of two restriction enzyme- a rare and a frequent cutter, e.g. *Eco*RI & *MSe*I; *Eco*RI & *PSt*I.
- ➢ Ligation of double stranded adapters ot cut ends of DNA fragments
- ➢ Pre-selective amplification
- ➢ Selective amplification
- ➢ Separation of amplified fragments
- ➢ Visualization using autoradiography

Strength:

- ✓ Stable amplification and high repeatability
- ✓ Can generate fingerprints of any DNA regardless of their origin, so highly suited for DNA fingerprinting
- ✓ Can act as bridge between genomic and physical maps
- ✓ Hyper-variability

Weakness:

- Time consuming procedure
- Required significant technical skills and financial resources
- Dominant in nature

iv) Simple Sequence Repeats (SSR):

These are also called as microsatellites or Short Tandem Repeats (STRs) or Simple Sequence Length Polymorphisms (SSLPs). These are variable tandem repeats ranging from 1-10 nucleotides, dispersed throughout the genomes of most eukaryotic organisms. These repeats may be di, tri, or tetra-repeats with repeat pattern being perfect, imperfect, compound or complex. If the repeat number is more than 10 nucleotides, they are referred as minisatelites. The conserved regions flanking these repeats are used to design the primers for PCR based amplification. The polymorphism discovered is due to variation in number of repeats in the individuals.

Strength:

- ✓ Abundant and uniformly distributed in the genome

- ✓ Hypervariable (large number of allele per locus)
- ✓ Co-dominant markers
- ✓ Highly reliable and reproducible assay
- ✓ Marker systems of choice in most of crop improvement programmes

Weakness:

- High developmental cost
- Often primers are species specific.
- Needs prior information of the sequence to design primers

v) Single Nucleotide Polymorphisms (SNPs):

The differences which are found at single nucleotide position are referred to as single nucleotide polymorphism or SNPs. This type of polymorphism results due to substitution, deletion or insertions. This is mostly biallelic and co-dominant in nature.

Strength:

- ✓ Abundant
- ✓ Sequence based robust markers
- ✓ Fast and high throughput
- ✓ Genome-wide distribution
- ✓ Possible to discovery of new alleles, hence useful in allele mining
- ✓ Co-dominant

Weakness:

- Most SNPs are biallelic and thus less informative than SSRs
- SNP assay is costly
- Sophisticated equipments and considerable efforts required

Each marker system has its own advantages and limitations. They vary with each other with respect to genomic abundance, locus specificity, polymorphism level, reproducibility, technical requirements and financial investments. Hence, due care has to taken before choosing marker system for different genetic studies.

Applications of molecular markers in crop improvement:

- ❖ Assessment of genetic diversity
- ❖ Differentiation and identification of varieties
- ❖ Construction of genetic maps
- ❖ Mapping and tagging of genes/QTLs
- ❖ Marker assisted selection
- ❖ Physical mapping
- ❖ Map-based gene cloning

References & Suggested Readings

1. Waqas Malik, Javaria Ashraf, Muhammad Zaffar Iqbal, Asif Ali Khan, Abdul Qayyum, Muhammad Ali Abid, Etrat Noor, Muhammad Qadir Ahmad, and GhulamHasan Abbasi (2014) Molecular Markers and Cotton Genetic Improvement:Current Status and Future Prospects. http://dx.doi.org/10.1155/2014/607091
2. Beckman, J. S. and Soller, M. 1983. Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping and costs. Theoretical and Applied Genetics 67:35-43.
3. Ganal, M. W., Altman, T., Röder, M. S. 2009. SNP identification in crop plants. Current Opinion in Plant Biology 12:211–217.
4. Gupta, P. K., Rustgi, S. and Mir, R. R. 2008. Array-based high-throughput DNA markers for crop improvement. Heredity 101: 5-18.
5. Mohan, M., Nair, S., Bhawat, A., Krishna, T. G., Yano, M., Bhatia, C. R. and Sasaki, T. 1997. Genome mapping, molecular markers and marker-assisted selection in crop plants. Molecular Breeding 3: 87-103.
6. Mueller, U. G. and Wolfenbarger, L. L. 1999. AFLP genotyping and fingerprinting. Trends in Ecology and Evolution 14: 389–394.
7. Paterson, A. H. 1996. Genome mapping in plants. Austin, TX: Landes, 330pp.
8. Peleman, J. and van der Voort, J. R. 2003. Breeding by Design. Trends in Plant Sciences 8: 330-334
9. Powell, W., Machray, G. and Provan, J. 1996. Polymorphism revealed by simple sequence repeats. Trends in Plant Sciences 1: 215-222.
10. Reiter, R., Phillips, R. L. and Vasil, I. K. (Eds.). 2001. DNA-Based Markers in Plants. Kluwer Academic, Dordrecht. pp. 9–29.

Construction of Linkage Map in Crop Plants

Santosh, H. B.[1]*, A. Manivannan[2] and V. N. Waghmare[1]
[1]ICAR-Central Institute for Cotton Research (CICR), Nagpur
[2]ICAR- Central Institute for Cotton Research, RS, Coimbatore
*Email ID: santoshgpb@gmail.com

Precise mapping of traits of economic interest is essential for application of molecular marker technology for crop improvement. High resolution genetic mapping requires construction of saturated linkage map which determines the relative distance between the marker(s) and trait of interest. This genetic distance is estimated based on the mean number of recombination events, involving a given chromatid, in that region per meiosis. A linkage map is a genetic map of a species or experimental population that shows the position of its known genes or genetic markers relative to each other in terms of recombination frequency, rather than a specific physical distance along each chromosome. First ever genetic map was developed by Prof. Steve D Tanksley in 1995 in tomato using RFLP markers. Development of appropriate mapping population is first step in construction of linkage map and genetic mapping of traits.

Steps involved in construction of genetic linkage map:

a) Selection of parents and development of appropriate mapping populations.

Parents selected should have high genetic divergence both at genotypic level (to be measured by molecular markers) and at the phenotypic level (assessed as diversity for traits of interest). Geographical origin and pedigree or source of the lines selected as parents should be given importance while choosing parents. Different mapping populations can be constructed by planned crosses for mapping both qualitative and quantitative traits. Mapping populations like F2, F2:3, backcross (BC) and near-isogenic lines (NILs) can be utilized for mapping mono/oligogenic traits while immortal populations like recombinant inbred lines (RILs) and doubled haploids (DH) are most preferred for precise mapping of quantitative trait loci (QTL) or polygenic traits. The type and size of mapping populations can exert an influence on the accuracy and economic significance of genetic maps. Larger mapping population is always better especially when the goal is high resolution mapping in specific genomic regions or mapping QTLs of minor effect. Immortal populations of large size (preferably more than 200 individuals), genotyped by co-dominant markers yield more precise and high resolution linkage maps.

b) Identification of informative markers between parents.

Different probe-enzyme combinations (in case of hybridization based markers) and primers (in case of PCR based markers) should be used for assaying the parents selected and polymorphic markers between parents should be identified.

c) Genotype the mapping populations using the informative markers.

The individuals of the mapping population developed should be precisely genotyped using genome-wide distributed markers polymorphic between the parents. The generated molecular profiles should be scored either manually or using software to determines the allelic profiles in the mapping populations.

d) Analysis of segregation pattern of markers in the populations.

A molecular marker is expected to follow Mendelian inheritance, which can be tested using chi square test. The expected ratio for co-dominant markers is 1:2:1 and 3:1 for dominant markers. Wherever there is a problem of segregation distortion (any significant deviation from expected Mendelian ratio), either eliminate those markers from the analysis or increase the log of odds (LOD) score.

e) Establishment of linkage relationship.

Genetic linkage map is developed based on the recombination frequency of the markers segregating in the mapping population. The distance between the markers is determined by the extent of recombination between markers; greater the recombination frequency farther the markers are and lower the frequency closer the markers are. The markers that are genetically linked (less than 50% of recombination frequency) form a linkage group. Possibility of linkage is tested using 2 point or 3 point test, later being more preferred as it will reveal the order of the markers apart from distance between the markers. Linkage between markers is usually calculated using odds ratios (i.e. the ratio of linkage versus no linkage). The ratio is expressed as logarithm of odds ratio (LOD). Higher LOD score, higher is the reliability of the map constructed. Threshold LOD values need to be determined depending upon the situation. Haldane and Kosambi mapping functions are used to convert recombination frequency into centi Morgans (cM). Distance on a linkage map is not directly related to the physical distance of DNA between genetic markers. Linkage map can be constructed using different statistical packages like, Linkage1, GMendel, MapMaker, MapManager, JoinMap etc. The accuracy of measuring the genetic distance and determining marker order is directly related to the number of individuals in the mapping population. How good or saturated is the map constructed will be indicated by map saturation indices viz.,, the number of linkage groups (in any species should be equal to haploid number of chromosomes), no/few gaps and no chromosome end extension.

f) Defining telomeric and centromeric regions and assigning linkage group to chromosomes.

Telomeric regions can be identified by techniques like Fluorescent In-Situ Hybridization (FISH) and use of telomeric sequences as probes. Centromeric regions can be mapped using Fluorescent In-Situ Hybridization (FISH) or using cytogenetic stocks, like Telotrisomy and Ditelotrisomy. Once saturated linkage map is constructed, and telomeric and centromeric regions are mapped, we need assign different linkage group constructed to different chromosome which is achieved with the help of known genetic marker or trisomic stocks or translocation stocks. In some crops, special cytogenetic stocks are available such as chromosome additional lines (brassica and wheat) and B-A translocation stocks (maize) which are helpful in assigning linkage groups to chromosomes of the species.

References and Suggested Reading

1. Guo et al. (2007) A Microsatellite-Based, Gene-Rich Linkage Map Reveals Genome Structure, Function and Evolution in Gossypium, Genetics, 176: 1, 527-541.
2. Kulwal PL (2018) Trait Mapping Approaches Through Linkage Mapping in Plants. Adv Biochem Eng Biotechnol.; 164:53-82. doi:10.1007/10_2017_49.
3. Li et al. (2017) High-density genetic linkage map construction by F2 populations and QTL analysis of early-maturity traits in upland cotton (*Gossypium hirsutum* L.) doi: 10.1371/journal.pone.0182918.
4. Ma et al. (2008) Simple Sequence Repeat Genetic Linkage Maps of A-genome Diploid Cotton (Gossypium arboreum) Journal of Integrative Plant Biology, 50 (4): 491–502
5. Mulualem T and Bekeko Z (2016) Advances in Quantitative Trait Loci, Mapping and Importance of Markers Assisted Selection in Plant Breeding Research, International Journal of Plant Breeding and Genetics, 10 (2): 58-68
6. Semagn K., Bjørnstad Å. and Ndjiondjop M. N. (2006) Principles, requirements and prospects of genetic mapping in plants, African Journal of Biotechnology, 5 (25), 2569-2587.

QTL Mapping in Crop Plants

Santosh, H. B.[1]*, A. Manivannan[2] and V. N. Waghmare[1]
[1]ICAR-Central Institute for Cotton Research (CICR), Nagpur
[2]ICAR- Central Institute for Cotton Research, RS, Coimbatore
*Email ID: santoshgpb@gmail.com

The term 'Quantitative Trait Locus (QTL)' was first coined by Gelderman (1975) which The term Quantitative Trait Locus (QTL) was first coined by Gelderman (1975) which defines QTL as "a region of the genome that is associated with an effect on a quantitative trait". Conceptually, a QTL can be a single gene, or it may be a cluster of linked genes that affect the trait. A quantitative trait exhibits continuous variation in its phenotypic expression. The trait expression may be controlled by many genes (a few hopefully of *large effect* and the others of *small effect*). Phenotypic expression of the trait is affected by genetic (*gene × gene interaction; epistasis*) and non-genetic factors [(*macro-* and *micro-environment variations* and *measurement errors* (*Noise*)] apart from their interaction with the environment (*gene × environment interaction*)**.** A genetic marker that tends to co-segregate with the trait is likely to be close to a QTL controlling that trait. In QTL mapping, we seek an association between marker alleles (genotypes) and trait values (phenotypes). The major purpose of QTL mapping is primarily to estimate the locations, number, magnitude of phenotypic effects, and mode of gene action, of individual genetic loci (QTLs) for its application in marker-assisted breeding. Assessing the relative contribution of a QTL to the total genetic variance for a specific trait (R^2), analysis of QTL x E effects and identification of parental lines contributing the favourable QTL allele are the other important components of QTL analysis. The fundamental challenge in QTL analysis is to disentangle the genetic signal at any individual locus from the noise. With Polymorphic markers, QTL analysis can be done with or without a genetic linkage map. A linkage map needs adequate polymorphic markers and allows mapping of QTL, estimating their number, phenotypic effects, and mode of gene action. Without a linkage map, only QTL-harbouring markers can be identified; QTL location, number, effect and mode of gene action cannot be determined. QTL analysis can be done using a *mapping population* (Linkage analysis) or a *natural/breeding population* (Association analysis). Classical analysis of QTLs employed statistical techniques which were based on means, variances and co-variances provided negligible information on what these genes were, where they were located, and how they controlled the trait. Major advancements in PCR-based marker systems, applied statistics and quantitative genetics revolutionised the QTL research.

Requirements

1. A suitable mapping population generated from phenotypically contrasting parents
2. A reasonably saturated linkage map based on molecular markers
3. Reliable phenotypic screening of mapping population
4. Powerful statistical packages to analyse genotypic and phenotypic information in combination.