Research and Reviews: Journal of Agriculture and Allied Sciences
e-ISSN: 2347-226X
p-ISSN: 2319-9857

Molecular Markers as a New Generation Demand in Cotton (Fabrics of Our life) Breeding Programme

Ravi Prakash Shukla*

Department of Plant molecular biology, CSIR- National Botanical Research Institute, Lucknow, India

Research Article

Received: 21/07/2021
Accepted: 04/08/2021
Published: 11/08/2021

***For correspondence:**
Ravi Prakash Shukla, Department Plant molecular biology, CSIR-National Botanical Research Institute, Lucknow, India
E-mail: gnanakumar12000@yahoo.com

Keywords: Mapping population; QTLs; Molecular markers; EST-SSR; SNP

ABSTRACT

Cotton (*Gossypium spp.*) is major cash crop of India and this country is the second largest cotton producer in the world after China. G. hirsutum and G. Barbadense two tetraploid species are majorly cultivated. In cotton many traits like fiber quality, drought tolerance, ball weight, ball number and yield are controlled by polygene or Quantitative trait locus. The development of meaningful molecular markers from different resources of cotton in the mapping population background to accelerate the QTL identification using different arithmetical programme is intense for QTL mapping. Current review provides an update on comparative studies and application of different molecular markers (RFLP, AFLP, RAPD, SSR, EST-SSR, SNP etc.) in cotton breeding programme.

INTRODUCTION

Cotton (*Gossypium*) is the most important fiber crop not only of India but of the entire world. It provides basic raw material for textile industry. India is a second largest cotton producer in the world after China. Although yield per hectare are still low compared to other major producers such as China and the United states. Yields differ depending on type of agro climatic conditions and seed used. The greatest diversity of wild cotton species is found in Mexico, followed by Australia and Africa. Cotton was independently domesticated in the Old and New Worlds. Gossypium comprises about 50 species of trees, shrubs and herbs [1]. Out of them, 45 are diploid (2n =2x=26) and 5 are allo-tetraploid (2n=4x=52). According to homologous pairing of chromosome during meiosis and karyotyping, diploid species (2n = 26) are classified into eight genomic groups *viz*; A, B, C, D, E, F, G, and K. All five allo-tetraploid species comprise both at as well as Dt sub genomes. Cotton usually is cultivated by four species; these include two New World allopolyploids G. hirsutum L. and G. barbadense L. (2n=52), and two Old World diploids G. herbaceum L., and G. arboreum L. (2n=26). All these four cultivated species were domesticated independently at different times in ancient history.

Generally plants have two type of traits, first trait that is controlled by a single gene and second that is controlled by many gene called polygenic trait and they are influenced by environments consequently they produces different phenotype. Plant breeders select the advantageous traits by looking at phenotype. While quantitative genetic provides the collective effect of the loci involved in polygenic trait but fails to address a particular traits. The partition of quantitative traits in to entity genetic section by assisting the molecular markers tightly linked to each trait. Finding of these DNA markers may help to plant breeders to select desirable genotype at plantlet stage, while the conventional breeding contains the constrain at this level for desirable genotype selection for plant breeder. The identification of polygenic trait can be inundating with different molecular markers [2]. Many researchers have been published a number of inter and intra-specific genetic linkage map in cotton by finding a number of DNA markers. Speeding up to construct genetic linkage map in cotton there are different marker technologies are available ie., RFLPs, AFLPs, RAPDs, SSRs, EST-SSR, and SNPs. Each markers have own pros and cons.

LITERATURE REVIEW

Molecular markers in cotton

The term restriction fragment length polymorphism or RFLP refers to differences between two or more samples of homologous DNA molecules arising from differing locations of restriction sites, and to a related laboratory technique by which these segments can be distinguished; RFLP analysis was the first DNA profiling technique inexpensive enough to see widespread application. RFLP is an easy, absolute and reproducible method and first molecular marker technique which was developed for plant breeding. This technique has been widely used to monitor differentiation in DNA sequence based based upon restriction sites among species. In lots of crop RFLP haven been engaged to know the additional information of relationship between parents,to the genes for different traits and, for the naming of measurable characteristic loci.
The first RFLP evalution in upland cotton were constructed a first detailed RFLP map to explore chromosome organization and evolution in cotton. A total 705 RFLP Loci were sorted in 41 linkage groups, covering 4,675 cM of cotton genome. The major strength of RFLP markers are co-dominance and high reproducibility. The main drawback of RFLP is low level of polymorphism. In addition, either sequence information (for PCR analysis) or probes are required, making it difficult and lengthy to develop markers in species lacking known molecular information.

Random Amplified Polymorphic DNA (RAPD)

It is a type of PCR reaction, but the segments of DNA that are amplified are random. The researchers performing RAPD creates several arbitrary short primers (8-12 nucleotides), then continue with the PCR using large template of genomic DNA, hoping that fragment will amplify. By resolving the resulting patterns, a profile can be draw from a RAPD reaction. Among different marker technology the RAPD is fast, simply assessable and regardless of sequence data for primer designing. A trispecific F2 mapping population in cotton was developed in Arkansas. Ninety F2 plants were derived from a cross between G. hirsutum and synthetic tetraploid cotton made of two diploid species G. arboreum (A2 genome) and G.trilobum (D8 genome). The 219 loci were assembled into 40 linkage groups with coverage of 3,855 cM. Linkage groups were assigned to chromosomes by use of diploid and aneuploid cotton lines. They detected 10 fiber related QTLs. The major advantages of RAPD markers are no requirement of DNA probe and sequence information. It involves no blotting or hybridization steps, hence, it is quick sample and efficient. The main drawback of RAPD is less reproducibility, dominant, Lack of prior knowledge on the identity of the amplification product.

Amplified Fragment Length Polymorphism (AFLP)

The AFLP technique was developed in the early 1990s by Keygene. AFLP is a PCR-based multi-locus fingerprinting technique used in genetic research and practice in genetic engineering which utilizes arbitrary primer based upon restriction enzyme sites sequences to amplify restriction fragment. AFLP polymorphism arises due to insertions or deletions within the amplified fragments. The first report of application of AFLP in cotton appeared in 1997. successfully detected AFLPs in cotton, which led to the application of this technique for genetic mapping The AFLP technique has been used as a powerful tool to estimate genetic diversity in major crops such as rice, wheat, maize, pearl millet, sunflower and peanut. Moreover, this technique is also useful for cultivar identification and evolutionary relationship studies .A Linkage map of cotton was developed using AFLP and RAPD markers. The main

strengths of the AFLP method include high frequency of genome-wide polymorphic markers, reproducibility due to high PCR annealing temperatures, and are relatively economical on a per marker basis compared to other arbitrary-primed PCR-based molecular marker systems. Like RAPDs, it does not require any prior sequence information and consequently is appropriate for analysis of germplasm, biodiversity and genetic relationship studies in a wide range of specie. Their major flaw of AFLPs is their dominant fashion of inheritance and the difficulty in identifying allelic variants at a specific locus although co-dominant AFLP markers have been found, however, in frequencies of 4-15% among all polymorphic AFLP markers.

Simple Sequence Repeat (SSR)

Microsatellites, or simple sequence repeats (SSRs) are a PCR-based genetic marker system. The "satellite" refers to the fact that eukaryotes have families of repetitive DNA sequences in their genomes. The repetitive DNA can be separated from total genomic DNA *via* centrifugation and this is called "satellite" DNA. Depending on the length of the repeated sequence, the sequences have been termed "satellites", "minisatellites", and/or "microsatellites", etc. Defining micro and mini can be complicated, so the term "Simple Sequence Repeat" (SSR) may be handier. The strategy for using SSRs as genetic markers is that the repeat region may vary in length between genotypes, but the DNA flanking the repeat is sufficiently conserved that the same primers will work in multiple genotypes. Thus, a SSR polymorphism between two varieties is due to the difference in the length of the repeat between the two conserved sequences. SSRs are excellent markers because of their locus identity, high PIC value, multi-allelism, and PCR basis. SSRs can be assayed on a number of platforms. Nevertheless, high development costs and practical challenges are limitations of microsatellites.

The first genetic map completely SSR based using intra-specific F2 populations included 86 loci covering 666.7 cM, approximately 14.8% of the total recombination length of the cotton genome. Recently, comparatively dense SSR based genetic maps utilizing RILs derived from intra-specific populations have been reported with genome coverage of 865 and 1024 cM. Comparatively dense genetic maps have been constructed in an inter-specific crosses of cotton using SSRs [3]. A map for double haploids from inter-specific cross between, *Gossypium hirsutum* L. acc. TM-1, and *Gossypium barbadense* L. cv. Hai-7124 consisting of 510 SSRs and 114 RAPDs. The 489 loci were assembled into 43 linkage groups with coverage of 3314 cM of the cotton genome. They utilized the monosomic and telo-disomic lines of *Gossypium hirsutum* in a TM-1 background for chromosome association. SSRs have been employed to develop a genetic map that exploited another BC1 population derived from the inter-specific cross [{(Guazuncho2 (*G. hirsutum*) × VH8- 4602 (*G. barbadense*)} × Guazuncho2], which covered 4400 cM of the cotton genome. With addition of the 233 new SSR loci, the map comprised 1160 loci and 5519 cM, and provided wide coverage of the genome of tetraploid cotton. SSR based linkage maps are also developed from Recombinant inbred line poppulation derived from inter-specific crosses. One hundred and eighty-three RILs generated from a cross between an Upland cotton genotype (TM-1) and a Pima cultivar (3-79), was surveyed with SSRs and complex sequence repeats (CRS).

DISCUSSION

Constructions of SSR-based genetic maps have also been exploited by the F2 mapping population using an inter-specific crosses. A total of 205 SSRs, 107 RAPDs and 437 SRAPs were used to genotype F2 population derived from a cross between Gossypium hirsutum cv. Handan 208 and Gossypium barbadense cv. Pima 90. The developed map allocates 566 loci into 41 linkage groups with at least three loci in each group. Twenty-eight linkage groups were assigned to corresponding chromosomes by SSR markers with known chromosome locations. The map covered 5141.8 cM with a mean inter-locus distance of 9.08 cM. This backbone map was improved by integrating 463 new loci resulting in development of a map having 1029 loci (625 SSRs, 58 RAPDs, 341 SRAPs, and 5 REMAPs) assembled into 26 chromosomes/linkage groups, totally covering 5,472.3 cM of the allotetraploid cotton genome [4]. Construction of the cotton genetic map subsequently led to identification of QTLs for several traits including fiber related characteristics, economic attributes leaf defoliation parameters morphological traits and chlorophyll contents. Other uses of cotton microsatellite markers include physical mapping of Rf1 fertility restorer gene, and tagging of genes related to root knot nematode resistance virticulum wilt resistance. Moreover, markers associated with fiber traits have also been used in marker assisted selection.

Expressed Sequence Tag-Simple Sequence Repeat (EST-SSR)

The beginning of the genomics age has resulted in the production of vast amounts of publicly available DNA sequence data, including large collections of expressed sequence tags (ESTs) from a variety of different taxa. Recent research has revealed that ESTs are a potentially rich source of SSRs that reveal polymorphisms not only within the source taxon, but in related taxa, as well. A number of EST-SSRs have been developed in different laboratory and other newly released SSRs were used to enrich in genetic linkage map.

Single Nucleotide Polymorphism (SNP)

Single Nucleotide Polymorphisms are highly abundant markers that allow fine mapping of traits. Thus, SNPs are most widely utilized technology for genetic mapping. Methods to characterize such polymorphism are also amenable to high-throughput analysis. SNPs generally occur throughout the genome, both euchromatin and heterochromatin region but the frequency of SNP is found more in heterochromatin region. When a SNP occurs within a gene, it may yield a change in the encoded protein which may itself cause a phenotypic variation but sometimes it may not cause variation in protein structure according to the wobble hypothesis. Another important property of SNPs is that they are co-dominant (both alleles can be discriminately genotyped) which helps to determine whether an individual is homozygous or heterozygous. SNPs offer the option of constructing dense genetic maps, map-based gene cloning, marker assisted selection, marker assisted breeding and hap map-based association studies. They developed about 1,000 SNPs and 300 InDels by re-sequencing the ESTs of 24 Upland cotton genotypes. About 200 of these SNPs were also mapped in the TM-1 × 3-79 genetic map. Recently, research in cotton SNP discovery has been accelerated and many SNP markers have been reported. In spite, the number of cotton SNP markers is still low as compared to other major crops such as maize, soybean, rice. In cotton only 39,862 SNPs are still submitted in public domain (dB SNP), out of which 38,782 derived from G. hirsutum and rest 1,080 are from G.

Marker-Assistrd Selection (MAS)

Marker assisted selection or marker aided selection (MAS) is an indirect selection process where a trait of interest is selected based on a marker (morphological, biochemical or DNA/RNA variation) linked to a trait of interest (e.g. productivity, disease resistance, abiotic stress tolerance, and quality), rather than on the trait itself.

Marker-assisted selection (MAS) gained a wide credit as a proficient approach for improving quantitatively inherited traits; however, there are hardly any examples of successful MAS for complex polygenic traits, such as yield and drought resistance. Marker-assisted selection being very dramatically successful for introgression and pyramiding of major genes, still many objections remain too unresolved, but it continues to be dominated by transgene introgression programs and to a few degree backcross alteration programs for simple traits [5]. MAS greatly enhance the productivity and effectiveness in plant breeding programs as compared to conventional breeding methods. From first to last quantitative genetics research in cotton potentially accelerating new molecular tools like microarrays, ESTs and proteomics are being introduced allowing the identification of many genes and QTLs in cotton. We conclude that MAS is a useful approach to enhance drought-adaptive traits in cotton, but complimentary recombination and selection are required to combine these traits with high yield potential.

Genetic diversity studies in cotton

The knowledge of genetic relationship among genotypes or different species helps to address the complexity available germplasm, to discover the differences in available genotype for useful conservation plans. Depiction of genetic similarity among genotypes is meaningful source to select the parental combination for maintaining the genetic diversity in fabrics of our life breeding programme. Success of breeding programme depends on the genetic diversity among and within the available germplasm which enable to breeder for parental sources that will reproduce diverse populations for selection.

CONCLUSION

Molecular marker is an emerging technology in many crops breeding programme. It is an efficient molecular resource in background of genetic mapping population and QTL mapping. The demand of next generation breeding programme in many crops including cotton (fabrics of our life) can be fulfilled and these molecular marker resources are very helpful in finding of important trait or Quantitative trait locus on particular chromosomes in many

crop. Acceleration in breeding programme by molecular marker is very meaningful and time consuming and in last but not least, without marker no imagination of any breeding program.

REFERENCES

1. Vos P, Hogers R, Bleeker M, et al. AFLP: A new technique for DNA fingerprinting. Nucl Acid Res. 1995;23:4407-4414.
2. Meredith WR. RFLP association with varietal origin and heterosis. In Proc. Beltwide Cotton Prod. Res Conf Nashv. 1992;27:6-10.
3. Reinisch AJ, Dong JM, Brubaker CL, et al. A detailed RFLP map of cotton, Gossypium hirsutum x Gossypium barbadense: Chromosome organization and evolution in a disomic polyploid genome. Genet. 1994;138: 829-847.
4. Vos P, Hogers R, Bleeker M, et al. AFLP: A new technique for DNA fingerprinting. Nucl Acid Res. 1995;23:4407-4414.
5. Lukonge EP, Labuschagne MT, Herselman L. Combining ability for yield and fibre characteristics in Tanzanian cotton germplasm. Euphytica. 2007;161:383-389.