Bulgarian Journal of Agricultural Science, 26 (No 4) 2020, 816–828

Molecular markers and their application in cotton breeding: progress and future perspectives

Plamen Sabev[1], Nely Valkova[1] and Elena G. Todorovska[2*]

[1]*Agricultural Academy, Institute of Field Crops, 6200 Chirpan, Bulgaria*
[2]*Agricultural Academy, AgroBioInstitute, 1164 Sofia, Bulgaria*
[*]*Corresponding authors:* e.g.todorovska@gmail.com; plamensabev2000@yahoo.com

Abstract

Sabev, P., Valkova, N. & Todorovska, E. G. (2020). Molecular markers and their application in cotton breeding: progress and future perspectives. *Bulg. J. Agric. Sci., 26 (4)*, 816–828

Cotton is the most important natural fiber plant worldwide and a valued cash crop. It is also a significant source of oil and protein. The conventional breeding for genetic improvement has its limitations due mostly to the lack of knowledge about polygenic traits such as yield productivity, fiber quality, resistance to diseases, and adverse growing conditions. The use of molecular markers is one of the significant developments in the field of genetics, allowing faster detection and exploitation of DNA polymorphisms. Recent developments, including the sequencing of the genome of *G. raimondii*, *G. arboreum*, *G. hirsutum,* and *G. barbadense* open new possibilities for selection of the best traits present in the genome of cotton. This article gives an overview of various DNA markers used in cotton. They play a crucial role in cotton improvement programs, like (a) Analysis of genetic diversity, (b) Construction of linkage maps, (c) QTL analysis of agronomic and fiber-related traits, (d) Marker-assisted selection (MAS). There is no molecular marker type developed up to now which can fulfill all the requirements of the plant breeders. The scientists should take a holistic approach and use all tools at their disposal to utilize the full genetic diversity present in *Gossypium* spp.

Keywords: cotton; genome; DNA markers; MAS
Abbreviations: Restriction Fragment Length Polymorphism (RFLP); Random Amplified Polymorphic DNA (RAPD); Amplified Fragment Length Polymorphism (AFLP); Inter Simple Sequence Repeats (ISSR); Sequence Related Amplified Polymorphism (SRAP); Single Nucleotide Polymorphism (SNP); Simple Sequence Repeats (SSR), Genome Wide Associations Study (GWAS)

Introduction

Cotton (*Gossypium* spp.) has a long history of cultivation and a great economic importance being the world's most cultivated fiber crop (Cuming et al., 2015). It is an important source of vegetable oil (7th after palm, soya, rape, sunflower, palm kernel, and peanut oils) and protein meals (3rd after soya and rapeseed meal). Cotton is cultivated by more than 80 countries in the world. EU cotton production is estimated at less than 300 000 t, which represents only 1% of the world's production. Currently, cotton is produced only by three EU Member States, namely Greece, as the main grower, with 80% of European cotton area, followed by Spain with 20%. In Greece, cotton production is accounting for more than 8 percent of total agricultural output. After a period when Bulgaria produced cotton on less than 1000 ha, currently the production increases again. There is a quota for 3340 ha according to the Common Agricultural Policy. Outside EU, Turkey is a big player in cotton business. Cotton is a major cash and industrial crop and plays a key role in Turkish economy. Turkish cotton area and production are predicted to increase, as good returns and government support encourages production.

So far, 52 different *Gossypium* species, including 7 tetraploid (AD) and 45 diploid, differentiated into 8 genomes (A-G and K), have been described. Recently, *G. ekmanianum* was restored as a 6th polyploid species (Grover et al., 2015), distinct from both *G. hirsutum* (AD1) and *G. barbadense* (AD2). New species from Australia (*G. anapoides*) (Stewart et al., 2015) and Wake atoll (*G. stephensii*) (Gallagher et al., 2017) were also described. The first belongs to K6 genome while the present analyses of hundreds of nuclear gene sequences and of cpDNA of the second ones indicates that it is a new (7th) polyploid cotton species. The cultivated cotton includes two diploid A genome species (*G. arboreum* L., A_2A_2 and *G. herbaceum* L., A_1A_1) and two allotetraploid AD genome species (*G. hirsutum* and *G. barbadense* (AADD) (2n = 4x = 52). *Gossypium hirsutum* L. is the most important cotton crop with 95% of the world's production while *G. arboreum* and *G. herbaceum* together share 2% on a global level. Cultivated tetraploid cotton evolved about 1–2 million years ago through hybridization of an A genome donor species (*G. herbaceum* and *G. arboreum*) with a D genome (*G. raimondii* and *G. gossipioides*) followed by polyploidization.

The progenitor allotetraploid "AD" diverged and gives rise to "AD" tetraploid species *G. hirsutum* L. (Upland or Mexican cotton) and *G. barbadense* L. (Sea Island or Egyptian cotton). After polyploidization due to breeding selection directed toward improvement of specific traits in cotton a reduction in genetic diversity in the cultivated cotton is observed. The depletion of the genome occurred perhaps through a bottleneck during the process of domestication (Shaheen et al., 2012).

Global demand for cotton products is expected to increase 102% from 2000 to 2030. This is likely to occur in a world where arable land is decreasing, water supplies are declining, and the impact of global climate change on production is uncertain. This demands that the rates of production gains globally accelerate without a plateau. Ideally, these gains will be achievable with reduced inputs and will be neutral to farm size. Current rates of genetic gain for lint yield under normal plant densities range from 7.1 to 8.7 kg ha^{-1} yr^{-1}. The majority of this genetic gain has arisen from conventional breeding and selection, and biotechnology. The aim of plant breeders is to improve economically superior varieties or to combine interesting traits present in different parental lines of cultivated species or their wild relatives with an objective to enhance yield and quality of field crops. However, conventional breeding cannot sustain maximum levels of

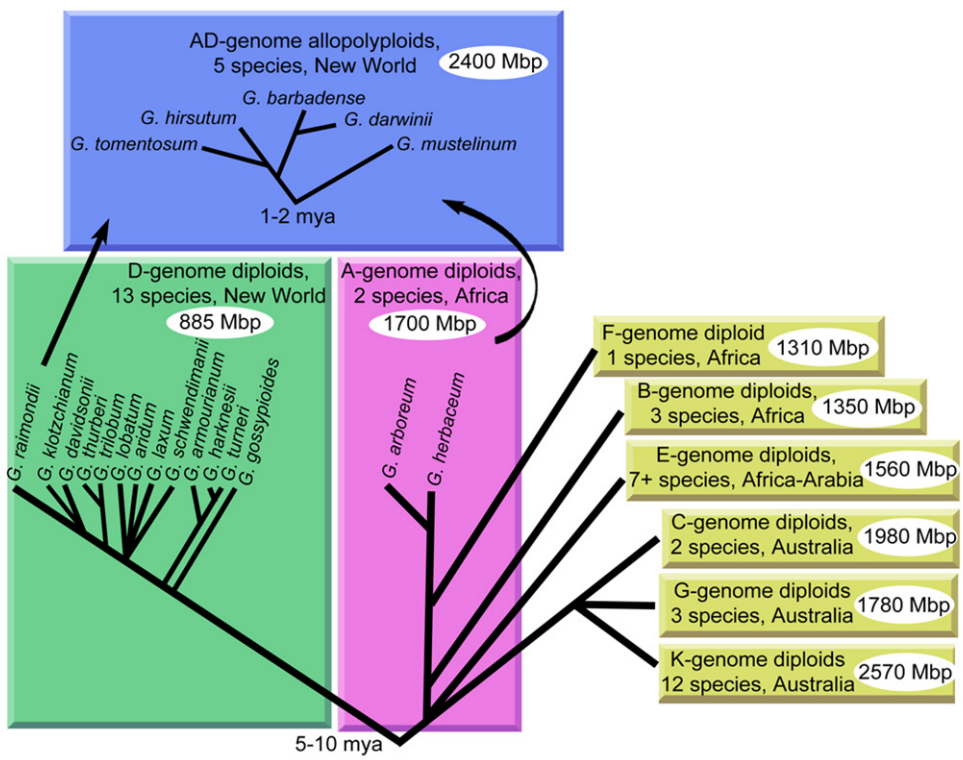

Fig. 1. Phylogeny and evolution of *Gossypium* species
(*Source*: Wendel Lab)

genetic gain without building upon supporting technologies. Fortunately, cotton genomic research has accelerated in recent years and this can promote enhanced cotton genetic improvement. This requires the adoption of a new approach in the selection of cotton varieties based on the use of molecular markers. The large and complex genomes of most of the cotton representatives require a large collection of DNA markers to achieve maximum genome coverage and utility in diverse germplasm.

The goal of this review is to follow the development of molecular marker technologies and their application in cotton improvement programs including the current MAS activities in cotton worldwide and in Bulgaria.

Molecular markers – potential and application in plant breeding

During the last two decades, the world has witnessed an exponential increase in plant sequences in databases which allowed not only molecular characterization of the genes but also the identification of molecular markers to facilitate crop improvement.

Genetic markers are one of the present developments in the field of genomics. A genetic marker is a fragment of DNA that is associated with a certain location on the chromosome or a gene whose phenotypic expression is easily recognizable and used to detect an individual. It can be described as a variation that can be observed. Markers show polymorphism, which may arise due to alteration of nucleotide or mutation in the genome loci and make it possible to identify genetic differences between individual organisms or species (Roychowdhury et al., 2013).

Genetic markers are divided into three groups:

(1) Morphological markers are phenotypic traits that can be monitored visually without specialized biochemical or molecular techniques. They are a very limited numbers.

(2) Biochemical markers are proteins produced because of gene expression, which can be separated by electrophoresis to identify the alleles. The most commonly used protein markers are isozymes that are variant forms of the same enzyme. Protein markers reveal differences in the gene sequence and function as co-dominant markers.

(3) Molecular or DNA markers represent sites of variation in DNA.

Traditionally, crop improvement has been achieved through the selection of observable phenotypes, representing the combined effect of all genes and the environment. This time-consuming effort is largely dependent on the performance of the selected candidates under certain environmental conditions. It is limited by the necessity that the phenotype has to be observable before the time when selection decisions have to be made or by its effectiveness in resolving a negative association between genes. Hence, plant breeders have a great interest in technologies that could make this procedure cheaper, less time-consuming, and efficient. Molecular markers are now a well-established technology that can be used in large breeding programs to complement the traditional breeding process. They are used in many different areas such as genotyping, genetic mapping, parental testing; detection of mutant genes related to hereditary diseases, cultivar identification, marker-assisted breeding, population history, epidemiology, and food safety, population studies. Important traits, such as yield, quality, and some forms of disease resistance are controlled by many genes and are known as polygenic traits and the environment influences many of them. This awareness about the extent of heredity about such traits on a molecular basis has shifted plant breeders to MAS. DNA markers linked to the QTL of interest increase the efficiency of breeding, decreasing costly and lengthy phenotypic selection. Transfer of required economic valuable characters from wild species to upland cotton without linkage drag is accomplished by MAS that is based on tracing of genomic regions in interspecific programs by molecular markers and QTLs (Abdurakhmonov et al., 2011). Through the increased numbers of next-generation sequencing, huge markers can be analyzed across the genome that allows genome-wide studies (Schuster, 2011). Although biometrical genetics provides the cumulative effects of the genetic loci involved in a polygenic trait, it fails to identify the locus/gene/s involved in a particular trait. If the quantitative traits are divided into individual, genetic components by finding a DNA marker closely linked to each trait, it would be easy to manipulate them efficiently and this would help to attain the desired results quickly and more precisely. These DNA markers will provide information to the plant breeders to select desirable plants directly on the basis of genotype in an early stage, instead of waiting for phenotype expression, where it is not possible through conventional breeding alone.

MAS has an advantage over conventional breeding, reviewed by many researchers (Kumpatla et al., 2012; Sundar et al., 2014). First, it is much quicker and, hence, much cheaper. Plant breeders use DNA markers for the selection of desirable traits on a molecular basis without a need to look at the phenotype, providing the basis for using the molecular assisted selection. Cotton is an important cash crop at a global level and is well suited to benefit from the latest developments.

DNA Markers in Cotton

DNA profiling in plants is mainly used for observing genetic diversity, germplasm maintenance, and determining

markers affiliated with required traits. Molecular markers are more authentic for fertility restoration than morphological markers in several lines of cotton. Genetic conservation is based on grip about the extent of genetic diversity prevailing in the germplasm. Molecular markers are easy to evolve due to the presence of enormous genomic databases and they are highly useful for plant breeders as these markers are a source of isolation, maintenance, detection of heredity, marker-assisted selection, and genomic profiling. Mishra et al. (2014) pointed out that the ideal DNA marker should be: highly polymorphic as it is compulsory for genetic studies; co-dominant, which allowed homozygotes and heterozygotes of diploids to be distinguished; detecting a relatively large number of alleles at many loci; phenotypic neutral; allowing cheap and fast assay; allowing reproducible and easy exchange of data among laboratories. The development of molecular markers is based on the cost of marker methodology, efficiency, and polymorphism. DNA markers are divided into three classes according to the method of their detection: (1) hybridization-based; (2) polymerase chain reaction (PCR)-based and (3) DNA sequence-based.

Hybridization Based DNA Markers

Restriction Fragment Length Polymorphism (RFLP)

Chronologically, RFLP is the first known type of molecular markers in plants which is used since 1975 for the detection of DNA sequence polymorphism with application in gene mapping. It belongs to the hybridization-based markers, which employs cloned DNA sequences to probe specific regions of the genome for variations that are seen as changes in the length of DNA fragments produced by digestion with restriction endonucleases. Dissimilarity in DNA sequences produces a loss, gain or alteration of the restriction site. Therefore, digestion of DNA with restriction enzymes produces fragments having a difference in number and size between populations and species. The main steps involve isolation of DNA, treatment with rare or frequently cutting enzymes, separation of restricted fragments by agarose gel electrophoresis, transfer of fragments to a nylon membrane, hybridization with DNA probe (gDNA segment or cDNA) and scoring of polymorphism by autoradiography.

RFLPs played a significant role in cotton genetics research until the years 2000–2005. They have been extensively used in diversity studies, variety identification, genetic map development, and QTL analysis, MAS for bacterial blight resistance (reviewed by Malik et al, 2014). The RFLP technique is, however, slow, expensive, and time-consuming. It requires a large amount of DNA, and the combined process of DNA fragmentation, electrophoresis, blotting, probe labeling, hybridization, washing, and autoradiography could take up to a month to complete. Therefore, nowadays it is limited in use, leading to the development of easier techniques known as PCR-based markers.

Polymerase chain reaction (PCR) DNA Markers

Randomly Amplified Polymorphic DNA (RAPD)

Random amplification of genomic DNA is the oldest PCR-based molecular marker technique. It utilizes random 10-mer primers for amplification of fragments without preliminary information about the DNA sequence using PCR. RAPD techniques have been used for many purposes by a number of researchers in cotton such as phylogenetic studies; genetic diversity; varietal identification (Chaudhary et al., 2010); intraspecific and interspecific differences and QTL analysis. RAPD marker (R-6592) linked to the male sterility gene has been identified in cotton (Lan et al., 1999). RAPD has many advantages over the RFLP technique such as nonradioactive detection, absence of prior sequence information, a small amount of genomic DNA, experimental simplicity, and no need for expensive equipment. The main disadvantage of RAPD markers is their poor reproducibility due to the influence of many factors like DNA concentration, PCR components i.e. DNA polymerase, and pipetting accuracy. Thus, the RAPD technique is notoriously laboratory dependent and needs carefully developed protocols to be reproducible.

Almost all RAPD markers are dominant-it is not possible to distinguish whether a DNA segment is amplified from a locus that is homozygous (1 copy) or heterozygous (2 copies). Co-dominant RAPD markers, observed as different-sized DNA segments amplified from the same locus, are rarely detected and they are due to short insertion/deletion in a specific locus. In some cases, mismatches between the primer and the template may result in the total absence of PCR product as well as in a merely decreased amount of the product. Thus, the RAPD results can be difficult to interpret.

Sequence Characterized Amplified Region (SCAR)

SCAR was derived from RAPD fragments. It is an improvement over RAPD by its capability to recognize just a single locus. The PCR reaction conditions are less manifested during amplification and they can be transformed into co-dominant markers. These markers are beneficial for genome mapping within closely related species, being co-dominant, and having the ability to evaluate pooled genomic libraries through PCR for map-based cloning. SCAR markers were used for the identification of disease and insect resistance and for the determination of fertility restoration in crops. The technology is cost-effective and produces highly polymorphic markers suitable in evaluat-

ing a large number of populations in cotton. In cotton, the SCAR 1920 marker for the major fiber strength QTL is well known (Guo et al., 2003). The main advantage of SCARs is that they are quick and easy to use. In addition, they showed high reproducibility and are locus-specific. Disadvantages are the need for sequence data to design PCR primers and that relatively few RAPD markers can be transformed into SCARs.

Cleaved Amplified Polymorphic Sequence (CAPS)

This is a technique for the analysis of genetic markers. It is an extension of the RFLP method, using PCR to quickly analyze the results. As RFLP, CAPS works on the principle that genetic differences between individuals can create or abolish endonuclease restriction sites and that these differences can be detected in the resulting DNA fragment length after enzyme digestion. Most CAPS markers are co-dominant and locus-specific, the genotypes are easily scored and interpreted. They are easily shared between laboratories.

These markers are useful in cotton and applicable especially in the characterization of germplasm, genetic diversity analysis for utilization in breeding programs, and genome mapping (Karaca & Ince, 2011). Their main weakness is that CAPS polymorphisms are harder to find because of the limited size of the amplified fragments (300–1800 bp) and the sequence data needed to design the primers.

AFLP (Amplified Fragment Length Polymorphism)

This technique combines the reliability of RFLP with the ease of RAPD. The process involves three simple steps: restriction of genomic DNA and ligation of oligonucleotide adaptors; pre- and selective amplification of restriction fragments; gel analysis of amplified fragments.

The main advantage of AFLP is its ability for "genome representation" to evaluate the representative DNA regions dispersed throughout the genome simultaneously. Generally, polymorphic fragments are detected as a present or absent making it a dominant marker system and rarely with a co-dominant nature. The technique can be automated and allows the simultaneous analysis of many genetic loci per experiments. AFLP produces more polymorphic loci per primer than RFLP, SSR or RAPD.

The benefits of AFLP are: reliable and reproducible profiles; absence of preliminary DNA sequence information for analysis; detection of a large number of polymorphic loci with a single primer combination on a single gel as compared to RFLPs and SSRs. Its main disadvantages are: the necessity of proprietary technology to score heterozygotes and ++ homozygotes; difficulty in developing of locus-specific markers from individual fragments; the need to use different kits adapted to the size of the genome being analyzed.

AFLP is an effective tool for assessment of genetic diversity, fingerprinting studies, and tagging of agronomic, seed and fiber quality traits in cotton. Due to their high abundance and random distribution throughout the genome AFLP markers have been effectively used for the construction of linkage maps and QTL analysis along with other markers (Rakshit et al., 2010; Ismail et al., 2015).

Sequence related amplified polymorphism (SRAP)

SRAP technique, developed in 2001, preferentially amplifies ORFs (open reading frames) through PCR using 2 different primers. The first one, forward primer consists of 17 nucleotides, in that 14 are rich in G and C at the 5'end and three selective ones at the 3'end. This primer amplifies mostly exonic regions. The second one is a reverse primer with 19 nucleotides, that contains a sequence of 16 nucleotides rich in A and T in the 5'end, and three selective nucleotides in the 3'end. This primer preferentially amplifies intronic and promoter regions. The polymorphisms fundamentally originate from the variation in the length of the introns, promoters and spacers, among both individuals and species. This technique combines simplicity, reliability, moderate throughput ratio and facilitates sequencing of selected bands. Further, it targets coding sequences in the genome and results in a moderate number of co-dominant markers. However, this technique does not utilize any prior sequence information, and the markers generated are randomly distributed across the genome. In cotton, SRAP is used along with other markers for saturation of the genome, germplasm evaluation and for construction of genetic linkage maps (Lin et al., 2005; Zhang et al., 2009).

Sequence Tagged Site (STS)

STS is a short sequence of DNA (up to 500 base pairs in length) that is seen to appear multiple times within an individual's genome. These sites are easily recognizable, usually appearing at least once in the DNA being analyzed and contain genetic polymorphisms making them sources of viable genetic markers (as they differ from other sequences). STS can be mapped within the genome and require a group of overlapping DNA fragments. PCR is generally used to produce a collection of DNA fragments. After overlapping fragments are created the map distance between STSs can be analyzed. In order to calculate it, researchers determine the frequency at which breaks between the two markers occur. STSs linked to desired characters were identified. Unlike arbitrary primers, STS relies on some pre-existing knowledge of the DNA sequence of the region. The DNA sequence of an STS may contain repetitive elements, but as long as the

sequences at both ends of the site are unique, complementary DNA primers can be synthesized and used for amplification and the specificity of the reaction is demonstrated by gel electrophoresis.

STS markers are simple to use, highly polymorphic, co-dominant, and suitable for high throughput sequencing. They tend to be more reproducible because longer primer sequences are used. Breeders used STS markers for developing of restorer parental lines for hybrid cotton (Feng et al., 2005). The investment in effort and cost needed to develop the specific primer pairs for each locus is their primary drawback.

SSR (Simple Sequence Repeat or Microsatellites)

Simple sequence repeats are polymorphic bands found in DNA that contain 1–6 bp repeating units. If the number of tandem repeats is higher than 10, then this marker shows a high level of inter and intra-specific polymorphism. Repetitive sequences are found scattered in both noncoding and coding regions of a genome. These markers are multiallelic, co-dominant, and intensively changing in comparison to other DNA markers. They are highly transferable across species. Due to their high level of polymorphism, SSRs are an important marker system in fingerprinting, analysis of genetic diversity, introducing novel genes into breeding materials from exotic germplasm, cultivar protection, screening of varieties, pedigree analysis, molecular mapping, and MAS. For PCR amplification of simple sequence repeats, oligonucleotides complementary to precise flanking fragments are used as primers. During replication tandem repeats produce SSR variants to copy choice recombination or dissimilarity occur in the specific nucleotide sequence due to unbalanced crossing-over.

Several methods have been pursued to develop SSR markers in cotton, including analysis of SSR-enriched small insert genomic DNA libraries (Ince et al., 2010; Kalia et al., 2011), SSR mining from ESTs, and large-insert BAC derivation by end sequence analysis. The information about SSR markers and their map position can be found in the Cotton Gene database (www.cottongen.org/find/mapped markers). Cotton Microsatellite database (CMD) now contains more than 10,000 publicly available microsatellites (http://www.cottonmarker.org).

In cotton, SSR markers are the desired type of markers as having a higher ability for use in phylogenetic studies and genetic diversity analysis in cotton and closely related species. They are used for genetic mapping and QTL analysis (John et al., 2012). In particular, researchers have used SSRs for refinement of fiber traits (Wang et al., 2017) and association mapping (Qin et al. 2015). Several genes for disease resistance have been tagged by SSRs, including resistance to cotton leaf curl virus, bacterial blight (Xiao, 2010), root-knot nematode (Jenkins et al., 2012), and verticillium wilt (Baytar et al., 2017).

Expressed Sequence Tags (EST-SSRs)

The last advances in genomic technologies have generated a large number of ESTs in cotton. They are short (200–800 bp in length), unedited, randomly selected single-pass sequence reads derived from cDNA libraries. ESTs may be used to identify gene transcripts, and are an instrument in gene discovery and in gene-sequence determination. Many of these are available in public databases, which give an opportunity to identify SSRs in ESTs by data mining. Transcribed regions of the DNA (EST-SSRs) are mostly maintained throughout the species compared to genomic SSRs from the untranslated regions and are having more substitution than genomic SSRs. The evolution of huge expressed sequence tags produces a valuable origin of PCR-based markers for targeting SSRs. Among divergent species in plants, about 1–5% of the expressed sequence tags have tandem repeats having acceptable length for the development of markers. The presence of simple sequence repeats in the EST-SSR markers was confirmed by cloning and sequencing of the amplified products. During gene expression, EST-SSRs have more chances of being functionally linked with phenotype variations than the genomic SSRs.

Many attempts were done for the development of EST-SSRs in cotton. They are quickly obtained by electronic sorting because of their presence in the expressed regions of the genome. They constitute about ½ of the SSR markers in cotton. EST-SSRs have been widely applied in genetic diversity studies (Zhang et al., 2011), linkage mapping (Lin et al., 2009), QTL mapping and marker-assisted breeding (Liu et al., 2012).

However, EST-SSRs exhibit a lower level of polymorphism than the conventional genic SSRs. Wang et al. (2015) utilized EST-SSRs for genetic mapping in cotton and observed that they are useful for the saturation of the allotetraploid genetic linkage map, genome evolution studies, and for comparative genome studies.

Inter-Simple Sequence Repeat (ISSR)

ISSR is a general term for a genome region between microsatellite loci. The complementary sequences to two neighboring microsatellites are used as PCR primers; the variable region between them gets amplified (Sharma et al., 2012). The limited length of amplification cycles during PCR prevents excessive replication of too long continuous DNA sequences, so the result will be a mix of a variety of amplified

DNA fragments which are generally short but varied much in length. ISSR markers usually show high polymorphism. It allows the detection of polymorphism in inter SSR loci using a primer (16–25 bp long) complementary to a single SSR and anneals at either the 3' or 5' end, that can be di-, tri-, tetra- or pentanucleotide. The technique combines many benefits of AFLPs and SSRs with the universality of RAPDs.

ISSRs have been used as valuable markers for revealing both inter and intraspecific variations in cotton. They provide an easy way for examining the polymorphic bands conversely to other molecular markers. The genetic diversity and agronomic traits in *G. hirsutum* and F_2 progenies, obtained by the cross between cotton cultivar Opal and 5 other cultivars, including Bulgar 539 (a.k.a Chirpan 539) have been assessed by Noormohammadi et al. (2013) using 20 ISSRs, 4 SSRs and 40 RAPDs. The result of the stepwise regression analysis showed a significant positive correlation between the marker ISSR-UBC807-1500bp and the second yield. The Bulgar 539 x Opal hybrid showed significantly higher fiber length compared to its parents, which makes this cultivar a new elite cotton genotype for future breeding. Interesting is the utilization of our cultivar Bulgar 539 which is characterized as very early and highly productive variety. Previously, Noormohammadi (2011) used another Bulgarian cultivar Bellizovar (a.k.a Beli Izvor) which is characterized by the highest lint percentage among the Bulgarian varieties, high adaptive capacity and was in production for many years.

Lately, Sethi et al. (2015) published the results from the assessment of genetic diversity among 95 *G. arboreum* genotypes belonging to the races '*indicum*', '*burmanicum*', '*sinense*', '*soudanense*', '*bengalense*', '*cernuum*' using ISSR markers. The selected 100 markers produced 397 bands, of which 368 were polymorphic. The study confirmed that the *G. arboreum* is an important germplasm resource for the improvement of tetraploid cotton in the breeding programs due to its desirable traits like the ability to withstand drought, salinity, and tolerance to several pests and diseases.

In spite of their usefulness ISSR method also has some practical limitations. Repetitive DNA is not easily analyzed by next-generation DNA sequencing methods, which struggles with homopolymeric tracts. Hence, microsatellites are normally analyzed by conventional PCR amplification and amplicon size determination, sometimes followed by Sanger DNA sequencing. The use of PCR means that microsatellite length analysis is prone to PCR limitations like any other PCR-amplified DNA locus.

Single Nucleotide Polymorphism (SNP)

Variations of a single nucleotide (A, T, G, or C) that occurs at a specific position of the genome are known as SNPs. These may occur in the noncoding, coding and intergenic regions of the genome, due to mutation, either transition or transversion, and deletion or insertion abnormality. They allow the detection of the genes due to the variations in the sequences of nucleotides and these are either non-synonymous or synonymous. Of particular interest are the non-synonymous SNPs, which may alter protein function and phenotype. The synonymous ones can alter mRNA splicing that also results in changes in the phenotype. The main advantage of SNP markers is their ease of data management along with their flexibility, speed, and cost-effectiveness. SNPs are highly secured markers being normally assigned and connected with morphological changes, which affect the phenotype directly. They are the easiest type of markers having minor heredity entities as an alone base and can produce a large number of markers. Bi-allelic SNP markers are straightforward to merge data across groups and create large databases of marker information since there are only 2 alleles per locus and different genotyping platforms will provide the same allele calls once proper data-entry has been performed.

These markers are important tools for linkage mapping, map-based cloning and MAS due to the high level of polymorphism. The co-dominant nature of SNPs makes them able to distinguish the heterozygous and homozygous alleles.

SNPs furnish fast and efficient genotyping of enormous populations by using next-generation sequencing methodology like Genotyping by sequencing (GBS). The method was described by Elshire et al. (2011). It is simple, quick, very specific, highly reproducible, and may allow access to regions of the genome that are inaccessible to sequence capture approaches. Its main advantages are-reduced sample handling, few PCR & purification steps, no DNA size fractionation, efficient barcoding system, simultaneous marker discovery & genotyping and it scales very well. The ultimate goal of functional genomics is to screen better plant types in crop improvement by sharing phenotypic information from phenotype to genotype. GBS will evolve first to capture more sequence variants and then to whole-genome resequencing. It can produce numerous SNPs and is appropriate for the gene pool maintenance, diversity analysis, genomic selection, gene mapping, and other plant improvement methodologies. This methodology has been used in a number of species, such as cotton and sorghum, following basic protocol with minor changes (Poland et al., 2012). GBS application is highly interwoven in cultivated cotton due to complicated allotetraploid genetic constitution and repetitive DNA (Li et al., 2014).

In *Gossypium*, SNPs were used for analyzing genetic diversity, phylogenetic analysis, construction of linkage maps,

QTL analysis and genetic mapping. Shaheen & Zafar (2010) examined SNPs among 2 accessions of *G. arboreum* in 30 conserved regions of ESTs and identified in a total of 27 SNPs in a region of 7804 bp having a frequency of one SNP/371 bp and one indel/1300 bp. Roche 454 pyrosequencing technique in allotetraploid cotton produced a large number of SNPs through reduced representation library sequencing (Byers et al., 2012). Wang et al. (2013) developed an intraspecific linkage map of *G. raimondii* using 18597 markers, including gSSRs, EST-SSRs, SRAPs, and SSCP-SNPs (single-stranded conformation polymorphism) which were used to identify QTLs related to yield and fiber quality as well as their component traits. Gore et al. (2014) constructed a linkage map and conducted a QTL analysis of 10 agronomic and fiber quality traits in a recombinant inbred mapping population of the cross between an upland cotton line and an elite *G. hirsutum* line with stabilized introgression from *G. barbadense*.

Two high-density genetic maps, one inter- and one intraspecific map were constructed by Hulse-Kemp et al. (2015) using CottonSNP63K. Then, 70K SNP chip based on *Illumina Infinium* genotyping assay was developed as a result of International Consortium-based Collaborations. This high-throughput genotyping assay represents a resource that could be used globally by public and private breeders, geneticists, and other researchers to enhance cotton genetic analysis, breeding, genome sequence assembly, and many other uses. The array and maps provide a foundation for genetic dissection of agronomical and economically important traits and crop improvement through genomics-assisted selection. It will also foster the positional cloning and genome assembly efforts. Lately, this high-throughput genotyping assay was used to search for functional SNP related to salt-tolerance in cotton (Wang et al., 2016). In 2017, 80K SNP array was developed and showed to be a powerful tool for genotyping *G. hirsutum* accessions and genome analysis (Cai et al, 2017). The fast-growing contribution of portable markers in cotton furnishes inexpensive way for gene isolation and linkage mapping for breeding cotton with desirable traits.

Association mapping

Utilization of the genetic diversity available in germplasm, genetic map construction, and QTL mapping for useful traits has been performed, using segregating populations and DNA marker techniques. However, it is difficult to identify closely linked markers for molecular breeding in a bi-parental population due to confined crossing-over. In addition, the density of polymorphism in a bi-parental population often is restricted as some minor QTLs are not detected. Therefore, a new methodology, named "association mapping" is used for QTL mapping, which relies on linkage disequilibrium (LD) among the loci and uses cultivars having distinct traits (Zhao et al., 2014). The idea is that traits that have entered a population recently will still be linked to the surrounding genetic sequence of the evolutionary ancestor, or in other words, will more often be found within a given haplotype, than outside of it. These associations of alleles between a marker locus and phenotypic locus must then be independently verified in order to show that they either (a) contribute to the trait of interest directly, or (b) are linked to/ in LD with QTLs that contribute to the trait of interest (Zhang et al., 2013). Association mapping overcomes many obstacles of traditional genetic mapping due to increased resolution generally to the locus and exploiting highly studied populations, which are having a genetic variation associated with the phenotypic variation. It is compulsory in LD mapping to characterize LD magnitude and pattern in a population under observation for the acquisition of desired objectives. The magnitude of relation, the extent of parental recombination, and linkage disequilibrium in gene pool permit the selection of most appropriate collection for association mapping. Seed yield, yield component, fiber quality, and disease resistance traits in cotton have been studied by association mapping. It enables scientists to assess the variation found in the germplasm resources (Abdurakhmonov et al., 2009).

The advantage of association mapping is that it can map quantitative traits with high resolution in a way that is statistically very powerful. However, it requires a large number of SNPs within the genome of the organism of interest and is therefore difficult to perform in species that have not been well studied or do not have well-annotated genomes.

Genome Wide Associations Study (GWAS)

With the discovery of SNP, it is now possible to study the whole GWAS with desired QTLs for developing highly saturated mapping populations in plants (Malik et al., 2014). GWAS or whole-genome association study is an examination of a genome-wide set of genetic variants in different plants or individuals to see if any variant is associated with a trait. This approach is known as phenotype-first, in which the organisms are classified first by their phenotype, as opposed to genotype-first. A sample of DNA is taken from each plant and millions of genetic variants can be read from it by SNP arrays. If one type of the variant (one allele) is more frequent in plants with the desired trait, the variant is said to be associated with it. The associated SNPs are then considered to mark a region of the genome that may influence the examined trait.

GWAS investigates the entire genome, in contrast to methods that specifically test a small number of pre-specified genetic regions. Hence, GWAS is a non-candidate-driven approach, in contrast to gene-specific candidate-driven studies.

While most published studies in crop species are candidate gene-based, genome-wide studies are on the increase. GWA studies identify SNPs and other variants in DNA associated with a trait, but they cannot on their own specify which genes are causal.

GWAS has been used for detecting favourable QTL alleles and candidate genes for lint percentage (Su et al., 2016) and for fiber quality traits (Sun et al., 2017) in upland cotton. The results indicated a significant association of 12 SNPs with lint percentage and 46 SNP with fiber quality traits, respectively.

Linkage Maps

Genetic linkage is the main exception to Mendel's Law of Independent Assortment. The linkage maps are used to distinguish between the chromosomes inherited from the parents. They demonstrate the location and the relative genetic distances in either side of markers across the chromosomes, which could be compared to signs along a roadway. The maps are useful in introgression research, examining genome structure and MAS in plant improvement studies due to close association with important agronomic traits (Li et al., 2017; Ronin et al., 2017; Fang et al., 2017). The genome of a crop is usually presented in the framework of a genetic linkage map. Such maps are used to locate or tag genes of interest, to facilitate MAS, and to enable map-based cloning. The use of MAS to improve the resistance to pests and diseases has become a choice for many breeders.

The regions in genomes having genes linked with a quantitative characteristic are named QTLs and QTL mapping is used for developing linkage maps and conducting QTL analysis. Principally the mapping is done by crossing-over that allows the analysis of genes and markers in the progeny. These characters are often of oligogenic inheritance in nature. Although, for some quality traits, few major QTLs or genes can account for a very high proportion of the phenotypic variation of the trait (Pham et al., 2012). Many required traits are examined at the same time by manipulating marker methodology which utilizes F2, recombinant inbred lines (RIL), back-cross populations, near-isogenic lines (NIL), and doubled haploids (DH). The recombination fractions are measured in centimorgans (cM) or map units (m.u.) during mapping analysis. A distance of 1 cM between two markers means that the markers are separated into different chromosomes on average once per 100 meioses. The investigation of many segregating markers produces a linkage map. Moreover, an additional marker's mapping may saturate the structure of maps. Marker types that produce multiple loci per primer combination like AFLPs are desired for increasing marker density. The selection of additional markers tagged to precise chromosomal regions may be observed by bulked-segregant analysis (BSA).

The first genetic map in cotton was constructed using RFLP markers and since then many genetic linkage maps have been developed and used in mapping QTL for lint yield, yield components, fiber quality. Other traits have been mapped and characterized using *G. hirsutum x G. barbadense* interspecific populations of F_2, BC, and RIL (Said et al., 2013).

However, few of them have been validated and applied in marker-assisted selection in cotton breeding programs. The comprehensive meta QTL analyses indicated QTL correspondence across researchers and populations was only modest. This finding suggested that for complex traits such as fiber quality, variation involved in a complex network of interacting genes and additional QTLs for the target traits remain to be discovered.`

While interspecific population genetic maps are already saturated to date, many genetic maps have also been constructed and used for QTL mapping with Upland cotton intraspecific populations (Tang et al., 2015: Liu et al., 2017; Zhang et al., 2017). Due to the narrow genetic background and low marker polymorphism, Upland cotton intraspecific linkage maps contain far fewer loci, leading to a lower resolution. Marker-rich linkage map for Upland cotton was constructed with a three-parent composite population, consisting of 978 SSR markers and spanning 4184.4 cM, with average marker distance 4.3 cM (Zhang et al., 2012).

Lately, Cao et al. (2014) investigated the first practical use of chromosome segment introgression lines (CSILs) for the transfer of fiber quality QTLs into upland cotton cultivars using SSR markers without severally affecting the economic traits. Adhikari et al. (2017) worked on the identification of an association between cotton fiber-quality traits and microsatellite markers. Recent advances in next-generation sequencing technologies have provided cost-effective platforms for direct detection of high-quality SNPs for genotyping of mapping populations. Genetic linkage maps and QTL analysis for fiber traits have been performed in diploid cotton (*G. arboreum* x *G. herbaceum*) (Badigannavar and Myers, 2015) and in upland cotton (G. *hirsutum*) (Tan et al., 2018; Zhang et al., 2019) with the aim to find stable QTL which is of high importance for fine mapping, identifying candidate genes, elaborating molecular mechanisms of fiber development, and application in cotton breeding programs by marker-assisted selection (MAS).

Application of molecular markers in Bulgarian cotton breeding

In Bulgaria, molecular markers were successfully applied to assess the level of genetic diversity in a wide variety

of crops of economic importance: cereals (reviewed by Todorovska et al., 2018), vegetables (Todorovska et al., 2014), etc. They were used to determine the level of frost tolerance in Bulgarian winter wheat (Todorovska et al., 2018) and the identification of interspecific hybrids between durum wheat and triticale (Bozhanova et al., 2014).

Breeding of cotton in Bulgaria has a long tradition and is mostly based on classical methods as hybridization and mutagenesis (Valkova & Dechev, 2004). As a result of successful selection in the last 20 years, many varieties were developed and released. This is a good prerequisite for the progress in the cotton production of our country. The development of early ripening varieties with high genetic potential for yield over 4500-5000 kg/ha and the selection of high fiber quality genotypes are the priority goals in the breeding programs in Bulgaria (Stoilova, 2011).

Genetic markers in cotton were successfully applied by for the development of a combined (RFLP-SSR-AFLP) genetic map of tetraploid cotton based on a *G. hirsutum* × *G. barbadense* backcross population (Bojinov & Lacape, 2003).

The progress in the cotton breeding program in Bulgaria was recently reviewed (Stoilova et al., 2014). The authors compared 11 new Bulgarian varieties with the standard ones: Chirpan-539 (for earliness and productivity) and Avanguard-264 (for fiber quality). The varieties were obtained from two different targeted programs of cotton breeding in Bulgaria: for productivity (Helius, Boyana and Vicky) and for fiber quality (Perla-267, Vega, Colorit, Darmi, Natalia, Rumi, Dorina, Nellina), and were created by the application of different breeding methods. Helius was developed by experimental mutagenesis (Valkova, 2009), Boyana and Vicky, by interspecific hybridization (Valkova & Bozhinov, 2010) while all others, by backcrossing of *G. hirsutum* L. x *G. barbadense* L. lines with the *G. hirsutum* L. varieties. Significant improvement of some agronomic characters over the standards was achieved. Perla-267 and Vicky exceeded the standard Chirpan-539 by 10.2 and 10.6% for seed yield while in fiber length (26.2-26.3 mm) they were close to the standard (26.4 mm average of three years). The best combination of length and fiber lint percent was found in the varieties Darmi, Vega and Boyana. The varieties Plovdiv, Philippopolis and Denitza are the latest examples of breeding for earliness and productivity.

Breeding studies in Bulgaria are focused on the development of new varieties and lines using diverse germplasm; maintenance and use of genetic resources; studies of heterosis effect based on male sterility; application of biotechnology approaches; drought tolerance; resistance to *Verticillium* and seed production. Genetic studies are focused on general and specific combining ability of lines and varieties and inheritance of most important agronomic traits; on the specific reaction of genotypes to the environmental conditions (genotype-environment interaction, stability, and plasticity); evaluation of genetic diversity and optimizing the breeding process using various statistical models.

However, a wider application of molecular marker in cotton breeding programs is needed to speed up and make the breeding process more cost-effective and efficient.

Conclusions

Marker-Assisted Selection in cotton is expected to have a huge impact on future breeding programs. Molecular markers offer a fast and relatively easy method of tracing genetic differences. Specific chromosome regions with important QTLs can be identified and used for efficient selection work. A major concern of cotton productivity decrease is the genetic uniformity of commonly used cotton cultivars that do not allow for making significant genetic improvement of yield-related traits, affected by biotic and abiotic stresses. However, this objective can be achieved by introgression and the use of modern molecular technologies in increasing the genetic gain of economic traits. DNA markers are the prominent type of genetic markers for MAS. SSRs have most of the desirable features and thus are the current marker of choice for many crops. The use of SNP markers in MAS programs has been growing faster and so the development of technologies and platforms for SNP discovery is an important task in many crops, including cotton. MAS has been successful for introgression and pyramiding major-effect genes, however, many challenges remain to be resolved before it can routinely provide added value for breeding very complex traits. MAS for qualitative traits appeared most successful after DNA fingerprinting while for quantitative characters, insect resistance genes, and genes controlling QTL for abiotic stress tolerance, the success is still limited. It is anticipated that the application of markers will remain restricted in these areas till the allele-specific markers are available and the cost of marker analysis is reduced significantly. Although there have been numerous QTL mapping studies for a wide range of traits in cotton, relatively few markers have actually been implemented in breeding programs for cotton improvement. The rate, scale, and scope of uptake of MAS in public crop breeding programs have continually lagged behind expectations. Many technical and logistical factors have hindered the speed and scope of MAS uptake. Steady progress and advancement in DNA markers will make it more attractive for molecular breeding and plant genetics and ultimately help in cotton improvement. The application of sequence-based genotyping for a whole range of diversity and genomic studies will have an important place well into the future.

References

Abdurakhmonov, I. Y., Buriev, Z. T. & Shermatov, S. E. (2011). Marker-assisted selection for complex fiber traits in cotton. In: *5th World Cotton Research Conference, Special session of ICGI,* Mumbai, India, 7-12.

Abdurakhmonov, I. Y., Saha, S., Jenkins, J. N., Buriev, Z. T., Shermatov, S. E., Scheffler, B. E., Pepper, A. E., John, Z.Y., Kohel, R .J., & Abdukarimov, A. (2009). Linkage disequilibrium based association mapping of fiber quality traits in *G. hirsutum* L. variety germplasm. *Genetica, 136(3)*, 401-417.

Adhikari, J., Das, S., Wang, Z., Khanal, S., Chandnani, R., Patel, J. D., Goff, V.H., Auckland, S., Rainville, L.K., Jones, D., & Paterson, A. H. (2017). Targeted identification of association between cotton fiber quality traits and microsatellite markers. *Euphytica, 213(3),* 65.

Badigannavar, A. & Myers, G. O. (2015). Construction of genetic linkage map and QTL analysis for fiber traits in diploid cotton (*Gossypium arboreum* x *Gossypium herbaceum*). *J. Cotton Sci, 19*, 15-26.

Baytar, A. A., Erdogan, O., Frary, A., Frary, A. & Doganlar, S. (2017). Molecular diversity and identification of alleles for *Verticillium* wilt resistance in elite cotton (*Gossypium hirsutum* L.) germplasm. *Euphytica, 213(2),* 31.

Bojinov, B. & Lacape, J. M. (2003). Molecular markers for DNA fingerprinting in cotton. In: *Cotton production for the new millennium: Proc. World Cotton Res. Conf., 3rd, Cape Town, South Africa,* 9-13.

Bozhanova, V., Todorovska, E., Hadzhiivanova, B., & Dechev, D. (2014). Obtaining of Interspecific Hybrids between Durum Wheat (2n=28) and Triticale (2n=42) and Molecular Evidence of Alien Introgressions in Advanced Backcross Line. *Türk Tarım ve Doğa Bilimleri, 6*(6), 868-874.

Byers, R. L., Harker, D. B., Yourstone, S. M., Maughan, P. J. & Udall, J. A. (2012). Development and mapping of SNP assays in allotetraploid cotton. *Theoretical and Applied Genetics, 124(7),* 1201-1214.

Cai, C., Zhu, G., Zhang, T. & Guo, W. (2017). High-density 80K SNP array is a powerful tool for genotyping *G. hirsutum* accessions and genome analysis. *BMC Genomics* **18,** 654.

Cao, Z., Wang, P., Zhu, X., Chen, H. & Zhang T. (2014). SSR marker assisted improvement of fiber qualities in *Gossypium hirsutum* using *Gossypium barbadense* introgression. *Theoretical and Applied Genetics, 127(3),* 587-594.

Chaudhary, L., Sindhu, A., Kumar, M., Kumar, R. & Saini, M. (2010). Estimation of genetic divergence among some cotton varieties by RAPD analysis. *J. Plant Breed. Crop Sci., 2(3),* 039-043.

Cuming, D. S., Altan, F., Akdemir, H., Tosun, M., Gurel, A. & Tanyolac, B. (2015). QTL analysis of fiber color and fiber quality in naturally green colored cotton (*Gossypium hirsutum* L.). *Turkish Journal of Field Crops, 20(1),* 49-58.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S. & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS one, 6*(5), e19379.

Fang, X., Liu, X., Wang, X., Wang, W., Liu, D., Zhang, J., Liu, D., Teng, Z., Tan, Z., Liu, F., & Zhang, F. (2017). Fine-mapping qFS07. 1 controlling fiber strength in upland cotton (*Gossypium hirsutum* L.). *Theoretical and Applied Genetics, 130(4),* 795-806.

Feng, C. D., Stewart, J. M. D., & Zhang, J. F. (2005). STS markers linked to the Rf 1 fertility restorer gene of cotton. *Theoretical and Applied Genetics*, *110(2),* 237-243.

Gallagher, J. P., Grover, C. E., Rex, K., Moran, M. & Wendel, J. F. (2017). A new species of cotton from Wake Atoll, *Gossypium stephensii* (Malvaceae). *Systematic Botany, 42(1),* 115-123.

Gore, M. A., Fang, D. D., Poland, J. A., Zhang, J., Percy, R. G., Cantrell, R. G., Thyssen, G., & Lipka, A.E. (2014). Linkage map construction and quantitative trait locus analysis of agronomic and fiber quality traits in cotton. *The Plant Genome*, *7(1).*

Grover, C. E., Zhu, X., Grupp, K. K., Jareczek, J. J., Gallagher, J. P., Szadkowski, E., Seijo, J. G. & Wendel, J. F. (2015). Molecular confirmation of species status for the allopolyploid cotton species, *Gossypium ekmanianum* Wittmack. *Genetic Resources and Crop Evolution*, *62(1),* 103-114.

Guo, W., Zhang, T., Shen, X., Yu, Z. J., & Kohel, R. J. (2003). Development of SCAR Marker Linked to a Major QTL for High Fiber Strength and Its Usage in Molecular-Marker Assisted Selection in Upland Cotton, *Crop Science, 43*(6), 2252-2256.

Hulse-Kemp, A. M., Lemm, J., Plieske, J., Ashrafi, H., Buyyarapu, R., Fang, D. D., Frelichowski, J., Giband, M., Hague, S., Hinze, L.L., & Kochan, K. J. (2015). Development of a 63K SNP array for cotton and high-density mapping of intraspecific and interspecific populations of *Gossypium* spp. *G3: Genes, Genomes, Genetics*, *5(6),* 1187-1209.

Ince, A. G., Karaca, M. & Onus, A. N. (2010). CAPS-microsatellites: use of CAPS method to convert non-polymorphic microsatellites into useful markers. *Molecular Breeding*, *25(3)*, 491-499**.**

Ismail, S. E., Adawy, S. S., El-Din, A. K. Y. G. & Hussein, E. H. (2015). Linkage map construction and detection of QTLs associated with earliness, fiber quality and yield in an interspecific cross between *Gossypium hirsutum* L. and *Gossypium barbadense* L. *International Journal*, *3(3),* 637-649.

Jenkins, J. N., McCarty, J. C., Wubben, M. J., Hayes, R., Gutierrez, O. A., Callahan, F. & Deng, D. (2012). SSR markers for marker assisted selection of root-knot nematode (*Meloidogyne incognita*) resistant plants in cotton (*Gossypium hirsutum* L). *Euphytica*, *183(1),* 49-54.

John, Z. Y., Kohel, R. J., Fang, D. D., Cho, J., Van Deynze, A., Ulloa, M., Hoffman, S.M., Pepper, A.E., Stelly, D.M., Jenkins, J.N. & Saha, S. (2012). A high-density simple sequence repeat and single nucleotide polymorphism genetic map of the tetraploid cotton genome. *G3: Genes, Genomes, Genetics*, *2(1),* 43-58.

Kalia, R. K., Rai, M. K., Kalia, S., Singh, R. & Dhawan, A. K. (2011). Microsatellite markers: an overview of the recent progress in plants, *Euphytica, 177(3):* 309-334.

Karaca, M. & Ince, A. G. (2011). New non-redundant microsatellite and CAPS-microsatellite markers for cotton (*Gossypium* L.). *Turkish Journal of Field Crops, 16(2),* 172-178.

Kumpatla, S. P., Buyyarapu, R., Abdurakhmonov, I. Y. & Mammadov, J. A. (2012). Genomics-assisted plant breeding in the 21st century: technological advances and progress. In: *Plant breeding*. Intechopen, Croatia, 131-184.

Lan, T. H., Cook, C. G. & Paterson, A. H. (1999). Identification of a RAPD marker linked to male fertility restoration gene in cotton (*Gossypium hirsutum* L.). *J. Agri. Genom., 4,* 1-5.

Li, F., Fan, G., Wang, K., Sun, F., Yuan, Y., Song, G., Li, Q., Ma, Z., Lu, C., Zou, C., & Chen, W. (2014). Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nature Genetics, 46(6),* 567.

Li, X., Wu, M., Liu, G., Pei, W., Zhai, H., Yu, J., Zhang, J. & Yu, S. (2017). Identification of candidate genes for fiber length quantitative trait loci through RNA-Seq and linkage and physical mapping in cotton. *BMC Genomics*, *18(1),* 427.

Liu, R., Wang, B., Guo, W., Qin, Y., Wang, L., Zhang, Y. & Zhang, T. (2012). Quantitative trait loci mapping for yield and its components by using two immortalized populations of a heterotic hybrid in *Gossypium hirsutum* L. *Molecular Breeding*, *29(2),* 297-311.

Liu, X., Teng, Z., Wang, J., Wu, T., Zhang, Z., Deng, X., Fang, X., Tan, Z., Ali, I., Liu, D., & Zhang, J. (2017). Enriching an intraspecific genetic map and identifying QTL for fiber quality and yield component traits across multiple environments in Upland cotton (*Gossypium hirsutum* L.). *Molecular Genetics and Genomics*, *292(6),* 1281-1306.

Lin, Z., He, D., Zhang, X., Nie, Y., Guo, X., Feng, C. & Stewart, J.M.C.D. (2005). Linkage map construction and mapping QTL for cotton fibre quality using SRAP, SSR and RAPD. *Plant Breeding, 124*, 180-187.

Lin, Z., Zhang, Y., Zhang, X. & Guo, X. (2009). A high-density integrative linkage map for *Gossypium hirsutum*. *Euphytica*, *166(1),* 35-45.

Malik, W., Ashraf, J., Iqbal, M. Z., Ali Khan, A., Qayyum, A., Ali Abid, M., Noor, E., Qadir Ahmad, M., & Hasan Abbasi, G. (2014). Molecular markers and cotton genetic improvement: current status and future prospects. *The Scientific World Journal*, *2014*.

Mishra, K. K., Fougat, R. S., Ballani, A., Thakur, V., Yachana, J. & Bora, M. (2014). Potential and application of molecular markers techniques for plant genome analysis. *Int. J. Pure App. Biosci.*, *2(1),* 169-188.

Noormohammadi, Z., Hasheminejad-Ahangarani, Y. F., Sheidai, M., Ghasemzadeh-Baraki, S. & Alishah, O. (2013). Genetic diversity analysis in Opal cotton hybrids based on SSR, ISSR, and RAPD markers. *Genetics and Molecular Research: GMR*, *12(1),* 256-269.

Noormohammadi, Z., Shojaei-Jesvaghani, F., Sheidai, M., Farahani, F. & Alishah, O. (2011). Inter simple sequence repeats (ISSR) and random amplified polymorphic DNA (RAPD) analyses of genetic diversity in Mehr cotton cultivar and its crossing progenies. *African Journal of Biotechnology*, *10(56),* 11839-11847.

Pham, A. T., Shannon, J. G. & Bilyeu, K. D. (2012). Combinations of mutant FAD2 and FAD3 genes to produce high oleic acid and low linolenic acid soybean oil. *Theoretical and Applied Genetics*, *125(3),* 503-515.

Poland, J. A., Brown, P. J., Sorrells, M. E. & Jannink, J. L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PloS one*, *7(2),* e32253.

Qin, H., Chen, M., Yi, X., Bie, S., Zhang, C., Zhang, Y., Lan, J., Meng, Y., Yuan, Y., & Jiao, C. (2015). Identification of associated SSR markers for yield component and fiber quality traits based on frame map and upland cotton collections. *PLoS One*, *10(1),* e0118073.

Rakshit, A., Rakshit, S., Singh, J., Chopra, S. K., Balyan, H. S., Gupta, P. K. & Bhat, S. R. (2010). Association of AFLP and SSR markers with agronomic and fibre quality traits in *Gossypium hirsutum* L. *Journal of Genetics*, *89(2),* 155-162.

Ronin, Y. I., Mester, D. I., Minkov, D. G., Akhunov, E. & Korol, A. B. (2017). Building ultra-high-density linkage maps based on efficient filtering of trustable markers. *Genetics*, *206(3),* 1285-1295.

Roychowdhury, R., Taoutaou, A., Hakeem, K. R., Gawwad, M. R. A. & Tah, J. (2013). Molecular marker-assisted technologies for crop improvement. *Crop Improvement in the Era of Climate Change*, 241-58.

Said, J. I., Lin, Z., Zhang, X., Song, M. & Zhang, J. (2013). A comprehensive meta QTL analysis for fiber quality, yield, yield related and morphological traits, drought tolerance, and disease resistance in tetraploid cotton. *BMC Genomics*, *14(1),* 776.

Schuster, I. (2011). Marker-assisted selection for quantitative traits. *Crop Breeding and Applied Biotechnology*, *11*(SPE), 50-55.

Sethi, K., Siwach, P., Verma, S. K. & Sihag, M. (2015). Assessing genetic diversity among *Gossypium arboreum* L. genotypes using ISSR markers. *International Journal of Pharma and Bio Sciences. 6(1),* 201-208.

Shaheen, T. & Zafar, Y. (2010). Detection of single nucleotide polymorphisms in the conserved ESTs regions of *Gossypium arboreum*. *Electronic Journal of Biotechnology*, *13(5),* 3-4.

Shaheen, T., Tabbasam, N., Iqbal, M. A., Ashraf, M., Zafar, Y. & Paterson, A. H. (2012). Cotton genetic resources. A review. *Agronomy for Sustainable Development*, *32(2),* 419-432.

Sharma, R., Joshi, A., Maloo, S. R. & Rajaman, G. (2012). Assessment of genetic finger printing using molecular marker in plants: A Review. *Sci Res Impact*, *1*, 29-36.

Stewart, J. M., Craven, L. A., Brubaker, C. & Wendel, J. F. (2015). *Gossypium anapoides* (Malvaceae), a new species from Western Australia. *Novon: A Journal for Botanical Nomenclature*, *23(4),* 447-451.

Stoilova, A. (2011). Breeding of Bulgarian cotton varieties with improved fiber quality. *Vavilov Journal of Genetics and Breeding*, *15*, 550-553.

Stoilova, A., Valkova, N., Hadzhiivanova, B., Koleva, M. & Nedyalkova, S. (2014). The cotton breeding in Bulgaria. *Türk Tarım ve Doğa Bilimleri Dergisi*, *1* (Özel Sayı-1), 992-999.

Su, J., Fan, S., Li, L., Wei, H., Wang, C., Wang, H., Song, M., Zhang, C., Gu, L., Zhao, S., & Mao, G. (2016). Detection of favorable QTL alleles and candidate genes for lint percentage by GWAS in Chinese upland cotton. *Frontiers in Plant Science*, *7*, 1576.

Sun, Z., Wang. X., Liu, Z., Gu, Q., Zhang, Y., Li, Z., Ke, H., Yang, J., Wu, J., Wu, L., Zhang, G., Zhang C., & Ma, Z.

(2017). Genome-wide association study discovered genetic variation and candidate genes of fibre quality traits in Gossypium hirsutum L. *Plant Biotechnol J., 15(8),* 982-996.

Sundar, R., Patil, S. S., Ranganath, H. M., Srivalli, P., Manjula, S. M. & Pranesh, K. J. (2014). Genetic diversity analysis in cotton (*Gossypium hirsutum* L.) based on morphological traits and microsatellite markers. *Int. J. Basic and Applied Biol.*, *1(1),* 19-22.

Tan, Z., Zhang, Z., Sun, X., Li, Q., Sun, Y., Yang, P., **Wang, W., Liu, X., Chen, C., Liu, D., Teng, Z., Guo, K., Zhang, J., Liu, D. & Zhang, Z.** (2018). Genetic map construction and fiber quality QTL mapping using the CottonSNP80K array in upland cotton. *Front. Plant Sci. 9*, 225.

Tang, S., Teng, Z., Zhai, T., Fang, X., Liu, F., Liu, D., Wang, S., Zhang, K., & Shao, Q. (2015). Construction of genetic map and QTL analysis of fiber quality traits for Upland cotton (*Gossypium hirsutum* L.). *Euphytica*, *201(2),* 195-213.

Todorovska, E., Abu-Mhadi, N., Christov, N., Bozhanova, V., Atanassov, A. (2018). Cereal genetics and genomics in Bulgaria - challenges and perspectives. *Comptes Rendus De L Academie Bulgare Des Sciences, 71(2)*, 143-160.

Todorovska, E., Ivanova, A., Ganeva, D., Pevicharova, G., Molle, E., Bojinov, B., Radkova, M., & Danailov, Z. (2014). Assessment of genetic variation in Bulgarian tomato (*Solanum lycopersicum* L.) genotypes, using fluorescent SSR genotyping platform. *Biotechnology & Biotechnological Equipment*, *28(1),* 68-76.

Valkova, N. (2009). Helius and Trakia new cotton varieties. *Field Crops Studies, 1,* 131-137 (Bg).

Valkova, N. & Bozhinov, M. (2010). Boyana cotton variety. *Field Crops Studies, VI-3*, 395-403 (Bg).

Valkova, N. & Dechev, D. 2004. Breeding evaluation of some traits and mutagenic cotton lines. *Field Crops Studies*, *I-3*, 402-406 (Bg).

Wang, B., Draye, X., Zhuang, Z., Zhang, Z., Liu, M., Lubbers, E. L., Jones, D., May, O. L., Paterson, A. H., & Chee, P. W. (2017). QTL analysis of cotton fiber length in advanced backcross populations derived from a cross between *Gossypium hirsutum* and *G. mustelinum*. *Theoretical and Applied Genetics*, *130(6),* 1297-1308.

Wang, Q., Fang, L., Chen, J., Hu, Y., Si, Z., Wang, S., Chang, L., Guo, W., & Zhang, T. (2015). Genome-wide mining, characterization, and development of microsatellite markers in Gossypium species. *Scientific Reports*, *5*, 10638.

Wang, X., Lu, X., Wang, J., Wang, D., Yin, Z., Fan, W., Wang, S., & Ye, W. (2016). Mining and analysis of SNP in response to salinity stress in upland cotton (*Gossypium hirsutum* L.). *PloS one*, *11(6),* e0158142.

Wang, Z., Zhang, D., Wang, X., Tan, X., Guo, H. & Paterson, A. H. (2013). A whole-genome DNA marker map for cotton based on the D-genome sequence of *Gossypium raimondii* L. *G3: Genes, Genomes, Genetics*, *3(10),* 1759-1767.

Xiao, J., Fang, D. D., Bhatti, M., Hendrix, B. & Cantrell, R. (2010). A SNP haplotype associated with a gene resistant to *Xanthomonas axonopodis* pv. *malvacearum* in upland cotton (*Gossypium hirsutum* L.). *Molecular Breeding*, *25(4),* 593-602.

Zhang, Z., Hu, M., Zhang, J., Liu, D., Zheng, J., Zhang, K., Wang, W., & Wan, Q. (2009). Construction of a comprehensive PCR-based marker linkage map and QTL mapping for fiber quality traits in upland cotton (*Gossypium hirsutum* L.) *Molecular Breed., 24,* 49-61

Zhang, K., Zhang, J., Ma, J., Tang, S., Liu, D., Teng, Z., Liu, D., & Zhang, Z. (2012). Genetic mapping and quantitative trait locus analysis of fiber quality traits using a three-parent composite population in upland cotton (*Gossypium hirsutum* L.). *Molecular Breeding*, *29(2),* 335-348.

Zhang, T., Qian, N., Zhu, X., Chen, H., Wang, S., Mei, H., & Zhang, Y. (2013). Variations and transmission of QTL alleles for yield and fiber qualities in upland cotton cultivars developed in China. *PLoS One*, *8*(2), e57220.

Zhang, Y., Wang, X. F., Li, Z. K., Zhang, G. Y. & Ma, Z. Y. (2011). Assessing genetic diversity of cotton cultivars using genomic and newly developed expressed sequence tag-derived microsatellite markers. *Genet. Mol. Res.*, *10*, 1462-1470.

Zhang, Z., Ge, Q., Liu, A., Li, J., Gong, J., Shang, H., Shi, Y., Chen, T., Wang, Y., Palanga, K.K., & Muhammad, J. (2017). Construction of a high-density genetic map and its application to QTL identification for fiber strength in Upland cotton. *Crop Science*, *57(2),* 774-788.

Zhang, K., Kuraparthy, V., Fang, H., Zhu, L., Sood, S., Jones, D.C. (2019). High-density linkage map construction and QTL analyses for fiber quality, yield and morphological traits using CottonSNP63K array in upland cotton (*Gossypium hirsutum* L.). *BMC Genomics 20(1)*, 889.

Zhao, Y., Wang, H., Chen, W. & Li, Y. (2014). Genetic structure, linkage disequilibrium and association mapping of Verticillium wilt resistance in elite cotton (*Gossypium hirsutum* L.) germplasm population. *PloS One*, *9(1),* e86308.

Received: February, 26, 2020; *Accepted:* March, 9, 2020; *Published:* August, 31, 2020