

REVIEW

Open Access

Role of SNPs in determining QTLs for major traits in cotton



MAJEED Sajid¹, RANA Iqrar Ahmad², ATIF Rana Muhammad¹, ALI Zulfiqar³, HINZE Lori⁴ and AZHAR Muhammad Tehseen^{1*}

Abstract

A single nucleotide polymorphism is the simplest form of genetic variation among individuals and can induce minor changes in phenotypic, physiological and biochemical characteristics. This polymorphism induces various mutations that alter the sequence of a gene which can lead to observed changes in amino acids. Several assays have been developed for identification and validation of these markers. Each method has its own advantages and disadvantages but genotyping by sequencing is the most common and most widely used assay. These markers are also associated with several desirable traits like yield, fibre quality, boll size and genes respond to biotic and abiotic stresses in cotton. Changes in yield related traits are of interest to plant breeders. Numerous quantitative trait loci with novel functions have been identified in cotton by using these markers. This information can be used for crop improvement through molecular breeding approaches. In this review, we discuss the identification of these markers and their effects on gene function of economically important traits in cotton.

Keywords: Abiotic stresses, Biotic stresses, Cotton, Earliness, Genotyping by sequencing

Background

Plant breeders are interested in genetic variations because these variations are the basis of phenotypic diversity. Many traits in plants arose due to genetic variations caused by mutation and/or recombination; those traits that were useful were 'fixed' by natural as well as artificial selection. With advances in technology, various methods have been developed by scientists to detect and analyze the minor genetic variations whose effects cannot be seen in the phenotypes (Jang et al. 2015). A base pair is the smallest unit of inheritance in an individual and when two or more individuals differ from each other based on a nucleotide then it is called a single nucleotide polymorphism (SNP). The identification of these minor variations was the initial challenging to plant scientists. The advent of next generation DNA sequencing technologies has solved this puzzle by being able to detect new functional SNPs associated with diverse traits. This whole genome sequence data serves as a reference for the identification of polymorphism due to SNPs among the individuals of the

same species (Xie et al. 2010). A lot of re-sequenced data is also available to identify the sequence diversity within crop plants. This data revealed whether changes in the genome within a species arose due to one or multiple factors (DePristo et al. 2011). Indeed the function of several genes has also been modified due to changes in a nucleotide which led to differences at the phenotypic level within plants of a species (Chung et al. 2013; Shi et al. 2015). Plant scientists have also reported several functional SNPs associated with phenotypic changes in various accessions of crop plants (Jang et al. 2015; Arruda et al. 2016). Several assays have been reported for genotyping in plants and most of these assays depend upon various molecular markers (Lateef 2015). SNP markers are the most abundant and robust ones for high throughput genotyping of plants. These markers can be found in all regions of a genome and a single gene may contain multiple SNPs (Rafalski 2002; Alkan et al. 2011). They play a significant role in determining phenotypic differences in plants, animals, humans and microbes (Moen et al. 2008; De Souza et al. 2010).

Identification of the location of a particular gene, measurement of distance among genes and their arrangement on the chromosome is called genetic

* Correspondence: tehsenazhar@gmail.com

¹Department of Plant Breeding and Genetics, University of Agriculture, Faisalabad, Pakistan

Full list of author information is available at the end of the article



mapping (Semagn et al. 2006). Genetic maps play an important role for the identification of quantitative trait loci (QTLs) (Ganal et al. 2009; Poland et al. 2012). The co-dominant, abundant and cost-effective nature of identifying SNPs made them ideal for construction of genetic maps in plant species. Genetic maps based on SNPs have been developed in several crop species such as cotton (Byers et al. 2012), rice (Xie et al. 2010), maize (Buckler et al. 2009), soybean (Akond et al. 2013) and *Brassica* (Li et al. 2009). Likewise, genome wide association study (GWAS) using SNP markers is a useful tool to develop genome wide haplotypes (Yano et al. 2016) and to detect natural diversity in cotton (Huang et al. 2017) and other crops (Aranzana et al. 2005; Yu and Buckler 2006; Poland and Rife 2012; Pasam et al. 2012). Identifying patterns among SNPs is a good method to study the evolution of a species at the genomic level to understand the history of a population as well as genetic variation among individuals and the role of selection pressure in inducing variation (Morin et al. 2004). SNPs also provide information about evolution of the modern genome by comparing the sequences of various species (Lu et al. 2013). Phylogenetic analysis of diploid cotton species using SNP markers revealed that A_1 and A_2 genomes are 98% similar (Shaheen et al. 2016).

Detection of SNPs in plants

Several techniques have been reported for the detection of SNPs in crop plants. Genotyping by sequencing (GBS) has been widely used for the identification of SNPs because of its low cost, rare chances of error and lower DNA purification requirement (Davey et al. 2011). The first step to identify SNPs from GBS is the isolation of genomic DNA. After quantification, the DNA is digested with a restriction enzyme. The choice of restriction enzyme is very important. Two restriction enzymes can be used for double digestion. Methylation sensitive restriction enzymes can also be used for analysis of methylated DNA. Digested DNA is then ligated with adaptors tagged by specific end sequences for polymerase chain reaction (PCR) amplification and sequencing. Various bioinformatic analyses are carried out on sequencing data in order to identify SNPs. These SNPs are further experimentally verified for their functional annotation (Elshire et al. 2011). A disadvantage of GBS is that some important regions of the genome may be missing from genomic libraries because the selected restriction enzymes did not cut in those regions. Another drawback of GBS is potential errors during sequencing (Kim et al. 2016).

The restriction-site associated DNA sequencing (RAD-seq) technique is used for discovery of SNPs when a reference genome is not available (Andrews et al. 2016). With this technique, a P_1 barcoded adapter is ligated to short DNA fragments generated after DNA

digestion with restriction enzymes. Adapter-ligated fragments of different samples are combined and DNA is sheared. Then, P_2 adapter primers are ligated to the DNA for amplification of these fragments and to produce sequencing libraries (Bergey et al. 2013). This technique is independent of a reference genome and relatively inexpensive. The degree of genome coverage can also be adjusted (Reitzel et al. 2013). This method requires high quality DNA and loss of sheared restriction sites may occur due to sequence polymorphism (Suchan et al. 2016). Another technique developed for large scale SNP based genotyping is specific locus amplified fragment sequencing (SLAF-seq). In this method, DNA sample is first digested with *MseI* and then digested with *AluI*. The resulting fragments are amplified by PCR, adapters are added and fragments are purified to obtain sequence libraries (Sun et al. 2013). This low cost method is useful for sequence based genotyping of large populations but it does not cover the whole genome (Ma et al. 2015). Chromatin immunoprecipitation followed by sequencing (ChIP-seq) is a sequencing tool that is used to investigate gene expression, i.e., transcription factors (Johnson et al. 2007). This tool has been characterized as robust because it profiles protein-DNA interaction *in vivo* on a genome-wide scale. It has enabled breakthroughs in transcriptional regulatory networks in *Saccharomyces cerevisiae* and human DNA regulatory sequences (Song et al. 2016). This protocol has great potential but is challenging to perform in plants due to necessary vigorous disruption of cell walls, presence of phenolic compounds and polysaccharides, and limited selection of quality antibodies that give a strong signal.

Reporting of SNPs/QTLs in cotton

Fibre quality and yield traits

Cotton is an important fibre and oilseed crop in tropical, sub-tropical and temperate regions of the world. It is widely grown on an area of 33.4 million hectares with production of 121.4 million bales annually (Johnson et al. 2018). Among 50 species of cotton, the allotetraploid species *Gossypium hirsutum* (also known as upland cotton) is the most widely grown (Sekmen et al. 2014). Cotton fibres and linters are the ultimate product of this crop that determine its price in an international market (Bradov et al. 1997). Staple length, strength, fineness and uniformity ratio are main parameters which are used to estimate fibre quality. Yield of seed cotton is a complex attribute that depends upon various parameters like boll weight, number of bolls per plant and lint percentage (Tang et al. 1996). Several SNPs and SNP-QTLs have been reported for yield and fibre related traits. Potential SNPs reported in cotton for all traits discussed here are summarized in Table 1. The cotton 63 K SNP

Table 1 Characterization of reported SNPs/QTLs in cotton for various traits of economic interest, stresses and plant architecture

No.	Cotton species	Source of genotyping	Genotyping method	No. of Experimental Years/Locations/Environments	Traits	No. of QTLs/ QTNs/ SNPs discovered	Chromosomal location of major SNPs/QTLs	Ref.
Yield and fibre quality traits								
1	<i>G. hirsutum</i>	188 F ₈ RILs	Cotton 63 K SNP array	2 Locations	Fibre quality and yield	71 QTLs, 16 Stable QTLs	05, 09, 10, 14, 19, 20	Li et al. 2016
2	<i>G. hirsutum</i>	277 F _{2:3} population	GBS	3 Years	Fibre quality and yield	5 178 SNPs, 110 QTLs	17, 26	Diouf et al. 2018
3	<i>G. hirsutum</i>	231 RILs population	Cotton SNP 80 K array	4 Locations, 4 Years	Fibre quality and yield	256 QTLs, 57 Stable QTLs	-	Liu et al. 2018
4	<i>G. hirsutum</i>	169 accessions	Cotton SNP 80 K array	2 Years, 2 Locations	Fibre quality traits	342 QTNs	A01, A06, A07, D01, D12, D13	Li et al. 2018a
5	<i>G. hirsutum</i>	201 accessions	Cotton 63 K SNP array	4 Years, 5 Environments, 3 Locations	Fibre quality and yield	23 254 SNPs	24	Handi et al. 2017
6	<i>G. hirsutum</i>	F _{2:3} lines	RAD-seq	1	Fibre quality and yield	27 QTLs, 9 366 SNPs	-	Wang et al. 2015b
7	<i>G. hirsutum</i>	178 RILs	RAD-seq	6 Years, 2 Locations	Fibre quality and yield	134 QTLs	-	Wang et al. 2015a
8	<i>G. hirsutum</i>	355 accessions	SLAF-seq	4 Environments	Fibre quality traits	33 SNPs	D07	Su et al. 2016b
9	<i>G. hirsutum</i>	180 RILs	SLAF-seq	3 Locations, 2 Years	Fibre quality	6 254 SNPs, 95 QTLs	-	Ali et al. 2018
10	<i>G. hirsutum</i>	196 F _{6:8} RILs	Cotton SNP 63 k array	5 Years, 6 Locations	Fibre strength	63 QTLs	-	Zhang et al. 2017
11	<i>G. hirsutum</i>	196 F _{6:8} RILs	SLAF-seq	11 Environments, 6 Locations	Fibre quality	37 QTLs	25	Zhang et al. 2015
12	<i>G. hirsutum</i>	180 RILs	Cotton SNP 80 K array	3 Environments	Fibre quality	12 116 SNPs, 104 QTLs	-	Tan et al. 2018
13	<i>G. hirsutum</i>	555 RILs and 11 parents	GBS	4 Years, 2 Locations	Fibre quality	6 071 SNPs, 86 QTLs	A07	Islam et al. 2016a
14	<i>G. hirsutum</i>	98 F _{5:7} RILs	GBS	2 Years, 2 Locations	Fibre quality and agronomic traits	412 SNPs, 28 QTLs	08, 23	Gore et al. 2014
15	<i>G. hirsutum</i>	419 accessions	GBS	12 Environments	Fibre related traits	3 665 030	A07, A10,	Ma et al. 2018

Table 1 Characterization of reported SNPs/QTLs in cotton for various traits of economic interest, stresses and plant architecture (Continued)

No.	Cotton species	Source of genotyping	Genotyping method	No. of Experimental Years/Locations/Environments	Traits	No. of QTLs/ QTNs/ SNPs discovered	Chromosomal location of major SNPs/QTLs	Ref.
16	<i>G. hirsutum</i>	196 F _{6,8} RILs	SLAF-seq	11 Environments, 6 Locations	boll weight	total SNPs, 7 383 unique SNPs	D03, D11	
17	<i>G. hirsutum</i>	375 F3 lines	Mapping by sequencing (MBS)	1	MD52ne genes (fibre strength and quality)	5 521 SNPs, 146 QTLs	A03, A04	Zhang et al. 2016b
18	<i>G. hirsutum</i>	1 genotype	Amplicon cloning and Sequencing	1	<i>GhMyb8</i> gene <i>GhMyb10</i> gene (Fibre quality)	27 SNPs, 4 QTLs	03, 14, 24	Islam et al. 2016b
19	<i>G. hirsutum</i>	Fibreless mutant line	Transcriptome sequencing	1	Fibre initiation	23 SNPs, 44 SNPs	–	Hsu et al. 2008
20	<i>G. hirsutum</i>	277 varieties	Eco-TILLING	3 Years, 3 Locations	Fibre quality and seed traits (<i>GhSus</i> genes)	28 610 SNPs	12, 26	Ma et al. 2016
21	<i>G. hirsutum</i>	503 accessions	Cotton 63 K SNP array	4 Locations, 2 Years	Agronomic traits	24 SNPs	16, 19, 25	Zeng et al. 2016
22	<i>G. hirsutum</i>	355 accessions	SLAF-seq	2 Years, 2 Locations	Lint percentage	324 SNPs, 160 QTLs	–	Huang et al. 2017
23	<i>G. barbadense</i>	2 genotypes, 185 F ₂ , 25 F ₁	Cotton 63 K SNP array	1	Fibre quality and yield	12 SNPs, 2 QTLs	02, 08	Su et al. 2016a
24	<i>G. arboreum</i>	230 accessions, 13 accessions	GBS	3 Locations	Agronomic traits	460 SNPs, 29 QTLs	05, 06, 23	Kumar et al. 2019
25	<i>G. hirsutum</i> × <i>G. barbadense</i>	146 Backcross inbred lines (BILs)	RNA-Seq	5 Environments	Fibre length	17 883 108 SNPs	–	Du et al. 2018
26	<i>G. hirsutum</i> × <i>G. tomentosum</i>	107 BC ₃ S ₅ introgression lines	SLAF-seq	3 Locations, 2 Years	Fibre quality and yield	703 SNPs, 4 QTLs	05, 11, 12, 21	Li et al. 2017c
27	<i>G. hirsutum</i>	2 lines, 1 line, 1 line, 1 line;	Amplicon cloning and Sequencing	1	<i>R2R3-MYB</i> transcription factors	3 157 SNPs, 74 QTLs	A01, A13, DO6	Keerio et al. 2018
	<i>G. tomentosum</i>					108 SNPs	–	An et al. 2008

Table 1 Characterization of reported SNPs/QTLs in cotton for various traits of economic interest, stresses and plant architecture (Continued)

No.	Cotton species	Source of genotyping	Genotyping method	No. of Experimental Years/Locations/Environments	Traits	No. of QTLs/QTNs/ SNPs discovered	Chromosomal location of major SNPs/QTLs	Ref.	
	<i>G. hirsutum</i> × <i>G. barbadense</i>	186 RILs							
	Biotic and abiotic stress tolerance								
28	<i>G. hirsutum</i>	8 varieties	Cotton SNP 63 K array	1	Salt tolerance	7 087 SNPs, 1 282 Salinity SNPs	A04, D01	Wang et al. 2016	
29	<i>G. hirsutum</i>	277 F _{2:3} population	GBS	3 Environments	Salt tolerance	5 178 SNPs, 66 QTLs	A02, D02, A01, A05, A03, A10, D01	Diouf et al. 2017	
30	<i>G. hirsutum</i>	97 F _{5:9} RILs	GBS	2 Environments	Drought and salt tolerance	165 QTLs	–	Abdelraheem et al. 2018	
31	<i>G. hirsutum</i> × <i>G. tomentosum</i>	BC ₂ F ₂ population	GBS	1	Drought tolerance	10 888 SNPs	23, 25	Magwanga et al. 2018	
32	<i>G. arboreum</i>	2 Genotypes,	Amplicon cloning and sequencing	1	MT-sHSP gene (Heat stress)	21 SNPs	–	Shaheen et al. 2009	
	<i>G. hirsutum</i>	1 genotype,							
	<i>G. herbaceum</i>	1 genotype,							
	<i>G. sturtianum</i>	1 genotype,							
	<i>G. aridum</i>	1 genotype,							
	<i>G. gossypoides</i>	1 genotype,							
	<i>G. laxum</i>	1 genotype,							
	<i>G. stocksii</i>	1 genotype							
33	<i>G. hirsutum</i>	318 accessions	GBS	3 Years, 3 Locations	<i>Verticillium</i> wilt resistance, Fibre quality and yield	1 871 401 SNPs, 119 QTLs	A02, A13, D08	Fang et al. 2017	
34	<i>G. hirsutum</i>	299 accessions	SLAF-seq	2 Years	<i>Verticillium</i> wilt resistance	85 630 SNPs, 17 unique SNPs	A06, A08	Li et al. 2017b	
35	<i>G. hirsutum</i>	120 accessions	Cotton SNP 63 K array	2 Environments	<i>Verticillium</i> wilt resistance	21 171 SNPs, 3 cluster QTLs, 2 Major QTLs	19	Zhao et al. 2017	
36	<i>G. hirsutum</i>	196 RILs	Cotton SNP 63 K array	4 Years	<i>Verticillium</i> wilt Resistance	119 QTLs	–	Palanga et al. 2017	
37	<i>G. hirsutum</i>	F4:5 population of 285 families	Amplicon cloning and Sequencing	1	<i>Xanthomonas axonopodis</i> resistance (Bacterial blight)	4 SNPs	14	Xiao et al. 2010	
38	<i>G. hirsutum</i>	364 F _{2:3} population	Amplicon cloning and Sequencing	1	Cotton blue disease	4 SNPs	10	Fang et al. 2010	

Table 1 Characterization of reported SNPs/QTLs in cotton for various traits of economic interest, stresses and plant architecture (Continued)

No.	Cotton species	Source of genotyping	Genotyping method	No. of Experimental Years/Locations/Environments	Traits	No. of QTLs/QTNs/ SNPs discovered	Chromosomal location of major SNPs/QTLs	Ref.
Earliness								
39	<i>G. hirsutum</i>	137 RILs	GBS	6 Years	Earliness	6 295 SNPs, 247 QTLs	A06, D03	Jia et al. 2016
40	<i>G. hirsutum</i>	355 accessions	SLAF-seq	4 Environments	Earliness	81 675 SNPs, 11 Favorable SNP alleles	D01, D03	Su et al. 2016c
41	<i>G. hirsutum</i>	169 accession	Cotton SNP 80 K array	2 Years, 2 Locations	Earliness	49 650 SNPs, 29 Significant SNPs	A06, A07, A08, D01, D02, D09	Li et al. 2018b
42	<i>G. hirsutum</i>	170 F _{2:3} population	GBS	1	Earliness	3 978 SNPs, 47 QTLs	D03	Li et al. 2017a
Plant architecture and related traits								
43	<i>G. hirsutum</i>	121 genotypes	GBS	3 Environments	Plant height and fruit spur branch number	2 620 639 SNPs, 11 QTLs	A08, D02	Wen et al. 2019
44	<i>G. hirsutum</i>	355 accessions	SLAF-seq	2 Locations, 3 Years	Plant architecture traits	93 250 SNPs, 22 novel SNPs	D03	Su et al. 2018
45	<i>G. hirsutum</i>	F ₂ population	dCAPS	1	Short fruiting branch gene	1 SNP locus	D07	Zhang et al. 2018a
46	<i>G. hirsutum</i>	F _{2:3} lines	GBS	1	Plant height, height of fruiting branch node and number of vegetative shoot	17 QTLs	03, 04, 05, 07, 09, 17, 19, 23, 25	Qi et al. 2017
47	<i>G. hirsutum</i> × <i>G. barbadense</i>	168 F ₂ population	GBS	1	Nulliplex-branch gene (<i>gb_nbl</i>)	42 SNPs	16(D07)	Chen et al. 2015
48	<i>G. hirsutum</i>	3 parental lines and their BC ₁ F ₁ population	Amplicon cloning and sequencing	1	Virescent leaf expression	4 SNPs	20	Zhang et al. 2018b
49	<i>G. barbadense</i>	3 genotypes	cDNA library sequencing	2 Locations	Diversity of leaf transcriptomes	> 10 000 SNPs	–	Kottapalli et al. 2016
50	<i>G. hirsutum</i>	395 genotypes	Cotton 63 K SNP array	1	Diversity analysis	38 822 SNPs	02, 13, 15	Hinze et al. 2017
51	<i>G. hirsutum</i>	11 genotypes and their C ₅ S ₆ RILs	GBS	1	Genetic diversity	5 617 SNPs	08, 09	Islam et al. 2015
52	<i>G. hirsutum</i>	2 accessions,	GR-PCR and KASP assays	1	Genetic diversity	151 000 putative SNPs	–	Byers et al. 2012
	<i>G. barbadense</i>	2 accessions						

Table 1 Characterization of reported SNPs/QTLs in cotton for various traits of economic interest, stresses and plant architecture (Continued)

No.	Cotton species	Source of genotyping	Genotyping method	No. of Experimental Years/Locations/Environments	Traits	No. of QTLs/QTNs/ SNPs discovered	Chromosomal location of major SNPs/QTLs	Ref.
53	<i>G. hirsutum</i>	2 genotypes,	GBS	1	Genetic diversity	25 529 SNPs	–	Logan-Young et al. 2015
	<i>G. barbadense</i>	2 genotypes,						
	<i>G. herbaceum</i>	1 genotype,						
	<i>G. raimondii</i>	1 genotype						
54	<i>G. hirsutum</i> × <i>G. barbadense</i>	59 interspecific F ₂ individuals	GBS	1	Structural variability	4 999 048 SNPs	–	Wang et al. 2015c
55	<i>G. hirsutum</i> × <i>G. barbadense</i>	186 RILs	Golden Gate assay	1	–	247 SNPs	–	John et al. 2012
56	<i>G. hirsutum</i>	440 accessions,	GBS	1	Haplotype distribution	10 129 SNPs	–	Reddy et al. 2017
	<i>G. barbadense</i>	218 accessions						

array was used to identify 71 QTLs for fibre quality traits strongly linked with SNP markers. These QTLs are comprised of seven pleiotropic QTL clusters, 19 e-QTLs, five hotspots and nine novel QTLs (Li et al. 2016). The linkage mapping, chromosomal localization and phylogenomic characterization of six *MYB* genes were carried out in four tetraploid cotton species via SNP markers. These *MYB* genes are actively involved in fibre development. The amplicon cloning and sequencing method of genotyping was used to detect 108 SNPs for these genes. It was determined that all six *MYB* genes evolved independently and exhibited significant variation in the D genome as compared with the A genome (An et al. 2008). Keerio and colleagues used 107 introgression lines derived from an interspecific cross of *G. hirsutum* and *G. tomentosum* for QTL mapping. They used the SLAF-seq method to obtain SNP markers. In this study, 74 QTLs and five clusters were found that were related to various fibre quality parameters (Keerio et al. 2018). Islam and co-workers have detected and validated 5 617 SNPs in upland cotton using GBS (Islam et al. 2015). These researchers have also reported 6 071 SNPs and 86 QTLs for the *GhRBB1_A07* gene. The experiment revealed the potential role of this gene in determining quality of cotton fibres. To identify this gene, they used a multi-parent advanced generation inter-cross (MAGIC) population which was developed through random mating of diverse *G. hirsutum* parents (Islam et al. 2016a).

More recently, 110 QTLs and five key genes namely *Gh_D12G0410*, *Gh_D12G0969*, *Gh_D12G0093*, *Gh_D12G0435* and *Gh_D03G0889* were found to be involved in fibre development in intraspecific crosses of *G. hirsutum*. These QTLs were detected through the GBS approach (Diouf et al. 2018). Another research group detected 28 QTLs related to fibre quality and agronomic parameters in a recombinant inbred mapping population using the GBS approach. They found seven QTLs for fibre strength while one QTL was detected for lint yield (Gore et al. 2014). Liu et al. used 231 recombinant inbred lines (RILs) and the Cotton SNP 80 K array to identify 122 QTLs for yield related traits and 134 QTLs for fibre quality parameters. Of these QTLs, 57 were detected in multiple environments and, therefore, were named as stable QTLs. The same group has also found 348 quantitative trait nucleotides (QTNs) with 74 stable QTNs for yield and fibre related traits (2018). The research group of Su has recognized 12 SNPs and 2 highly stable QTLs for lint percentage through a GWAS of 355 accessions. They used the SLAF-seq method for genotyping these cotton lines. These SNPs could provide a source to improve lint yield through molecular breeding (Su et al. 2016a). In another study, researchers have discovered 37 QTLs on chromosome 25 in a RIL population of upland cotton using the SLAF-seq method.

These QTLs were related to various fibre quality attributes (Zhang et al. 2015). In a separate report, Zhang found 63 QTLs for fibre strength, and these QTL were highly stable in nature. The researchers have used the Cotton SNP 63 K array for genotyping. This chip contains SNPs from several cotton species including *G. hirsutum*, *G. barbadense*, *G. tomentosum*, *G. mustelinum*, *G. armourianum* and *G. longicalyx* (Hulse-Kemp et al. 2015; Zhang et al. 2017). SNPs were also used to construct a genetic linkage map through the SLAF-seq approach and identify QTLs for boll weight. One hundred forty-six QTLs were found in 11 environments, and 16 of these QTLs were classified as stable QTLs because they were detected in more than three environments (Zhang et al. 2016b). Resequencing of 419 upland cotton accessions led to the discovery of 3 665 030 SNPs. These accessions were phenotyped for 13 fibre related traits in 12 different environments. GWAS revealed the association of 7 383 unique SNPs and 4 820 candidate genes for these traits (Ma et al. 2018).

Biotic and abiotic stress tolerance

The cotton plant faces various stresses during its life cycle that limit the productivity of the crop around the world. A single base pair difference between genotypes may be the underlying reason for a differential response to environmental stresses. Many studies have been conducted to evaluate whether genomic information can be used to identify SNPs and QTLs related to biotic and abiotic stress tolerance. The GBS method has been exploited to construct a high density genetic map with 10 888 SNPs from segregating populations of an interspecific cross (*G. hirsutum* × *G. tomentosum*) to detect QTLs related to drought tolerance. Thirty-four thousand four hundred two (34 402) and 32 032 genes were also mined within the Dt and At sub-genomes, respectively, to understand the genetics of drought tolerance (Magwanga et al. 2018). Abdelraheem et al. mapped QTLs for drought and salt tolerance using an RIL population derived from a cross of two diverse parental lines. A total of 165 QTLs were discovered through the GBS approach in this study, with 15 QTLs associated with tolerance to salinity and drought stresses common to two environments, i.e., greenhouse and field conditions (2018). Likewise, a high-density linkage map was also constructed using a segregating population of an intra-specific cross between salt tolerant and salt susceptible genotypes. A total of 66 QTLs and 5 178 SNP markers were identified through GBS for 10 salinity tolerance related traits in three different environments. Out of these QTLs, 14 were designated as stable due to their presence in more than one environment. Nine and five stable QTLs were located in the Dt and At sub-genomes, respectively, and 12 key genes were found to be involved in conferring

salinity resistance at the seedling stage (Diouf et al. 2017). In another experiment, Wang et al. used salt tolerant and susceptible genotypes for mining SNPs using the Cotton 63 K SNP array. A total of 7 087 SNPs were mined, out of which 1 282 were highly related to salinity tolerance in cotton (2016). Beside salinity and drought, another major abiotic stress is high temperature, but the SNPs related to this stress are yet to be explored in cotton. Previously, 21 SNPs were reported for the mitochondrial small heat shock protein gene (*MT-sHSP*). These SNPs were identified through PCR amplification and sequencing of this gene derived from several cotton species (Shaheen et al. 2009).

Among biotic stresses, *Verticillium* wilt is one of the major threats to cotton production in the USA, China and Turkey (Baytar et al. 2017). This disease causes significant reduction in yield, and the pathogen can survive for several years in the soil (Zhang et al. 2016a). GWAS revealed 17 SNPs related to *Verticillium* wilt resistance through the SLAF-seq method of genotyping. These SNPs were stable in three different environments. QTL analysis also revealed that *CG02* (a disease resistance protein belonging to the *TIR-NBS-LRR* class) seems to be responsible for resistance to *Verticillium dahlia* (Li et al. 2017b). Likewise, Zhao et al. used the Cotton SNP 63 K array to detect SNPs and QTLs related to this disease in two different environments. The results revealed the presence of 21 171 SNPs across 120 accessions of *G. hirsutum*. Three clustered QTLs, two major QTLs, 12 functional genes and six mRNAs conferring resistance against *Verticillium* were also detected (2017). In another research report, genomic analysis of many accessions through GBS revealed three trait loci involved in *Verticillium* wilt resistance. A candidate gene (*Gh_D06G0687*) was also reported that conferred resistance to this pathogen by encoding an *NB-ARC* domain (Fang et al. 2017). Cotton blue disease is one of the major diseases of cotton in Brazil, and it is transmitted through aphids (Silva et al. 2008). Haplotype mapping of a large segregating population through amplicon cloning and sequencing using specific SSR primers revealed that resistance was conferred by four SNPs (Fang et al. 2010). Another four SNP markers were discovered through haplotype mapping that were highly associated with resistance to bacterial blight disease (*Xanthomonas axonopodis* pv. *Malvacearum*) (Xiao et al. 2010). Aside from these diseases, the productivity of cotton is also affected by cotton leaf curl virus, root rot and cotton mosaic virus. Moreover, a huge number of pest insects are associated with this crop, but no SNPs linked to these biotic stresses have been reported in the literature to our knowledge. Therefore, it is

important for molecular plant breeders to explore SNPs related to these biological threats in order to understand the basis of genetic resistance.

Earliness

Early maturity is an important feature which is essential if growing more than one crop per year or to escape from late season environmental stresses. An early maturing genotype also requires less irrigation as well as less fertilizer and chemical inputs (Bednarz and Nichols 2005; Cober et al. 2010; Akter et al. 2019). One study was conducted to detect SNPs related to early maturity in upland cotton using 137 RILs. Sequence based genotyping revealed that 6 295 SNPs and 247 QTLs were associated with six morphological traits related to earliness. These QTLs were deemed highly stable due to their identification in six consecutive years, i.e., 2010 to 2015 (Jia et al. 2016). In another project, the SLAF-seq genotyping strategy was used to identify SNPs related to six earliness linked traits from 355 *G. hirsutum* accessions grown in four different environments. A total of 81 675 SNPs and 11 highly favorable SNP alleles were discovered. GWAS also revealed a potential candidate gene (*CotAD_01947*) that was associated with early maturity (Su et al. 2016c). More recently, a GWAS was conducted to identify SNPs and genes associated with four earliness related traits. A total of 49 650 SNPs were discovered using the cotton SNP 80 K array, and 29 SNPs were highly associated with early maturity. In addition, two potential candidate genes (*Gh_D01G0340* and *Gh_D01G0341*) were also related to earliness (Li et al. 2018b). Likewise, the GBS method has been used to construct a high-density genetic linkage map to discover QTLs related to this trait. The linkage map was comprised of 3 978 SNPs, and 47 QTLs were detected. These QTLs were associated with six earliness qualities. A study of an early maturing cultivar revealed two highly expressed potential candidate genes (i.e., *Gh_D03G0885* and *Gh_D03G0922*) (Li et al. 2017a).

Plant architecture and other important traits

A combination of traits are desirable to increase productivity of the cotton crop. Plant architecture is an important factor that determines suitability of cotton genotypes for mechanical picking and as well as to improve yield (Song and Zhang 2009). This complex multigenic trait has been given less importance in cotton as comparing with wheat and rice where deployment of dwarfing genes led to the Green Revolution. To investigate the genetic basis of plant architecture, a GWAS experiment was conducted with 121 upland cotton genotypes. The researchers identified 2

620 639 SNPs, 11 QTLs and 5 candidate genes for two plant architecture traits, i.e., fruit spur branch number and plant height. The cotton accessions were genotyped with the whole genome resequencing approach and phenotyped in multiple environments (Wen et al. 2019). In another study, 93 250 SNPs for five plant architecture traits were found in 355 Chinese upland cotton accessions using the SLAF-Seq method. GWAS revealed 22 highly associated SNPs and 21 candidate genes for these traits (Su et al. 2018). Molecular analysis of the short fruiting branch gene was carried out in an F₂ population between two parents, one with short fruiting branches and the other with long fruiting branches. One SNP locus (*SNP_GH1570*) was found to be highly associated with short fruiting branches when using derived cleaved amplified polymorphic sequences (dCAPS). It was concluded that this SNP marker was useful for selection of cotton plants with short fruiting branches (Zhang et al. 2018a). A separate study revealed the presence of 17 QTLs associated with plant height, height of fruiting branch node and number of vegetative shoots. These QTLs were located on nine different chromosomes and were detected through the GBS method (Qi et al. 2017).

A *nulliplex-branch* mutant was developed to explore the position of flowers on the cotton plant. This mutant line exhibits flowers which arise directly from leaf axils on the main stem, without a fruiting branch, i.e., monopodial and sympodial branches. This trait is desirable so planting densities can be increased without using chemicals to regulate plant growth (Du et al. 1996). To discover the molecular basis of the *nulliplex-branch* mutant, a genetic map was constructed from a *G. hirsutum* by *G. barbadense* interspecific population. The map was comprised of 11 805 SNP markers which were identified through next generation sequencing. The analysis revealed that 42 SNPs were associated with *gb_nb1*, a recessive gene that controls the *nulliplex-branch* trait (Chen et al. 2015). Virescent leaves in cotton are characterized by their yellowish appearance at early stages of plant growth. This abnormality is due to a recessive gene, *v₁*. Sequence analysis of wild and mutant alleles showed the differences in four SNPs at sequence positions 426, 450, 709 and 1 082. It was further revealed that the SNP at position 1 082 caused a point mutation that resulted in synthesis of arginine instead of lysine in mutant polypeptides (Zhang et al. 2018b). In another study, genetic diversity for leaf transcriptomes was identified in *G. barbadense*. Through a cDNA library sequencing technique, researchers have found more than 10 000 SNPs associated with various traits in three Egyptian cotton cultivars (Kottapalli et al. 2016). Likewise, many SNP markers were also identified using the GBS

approach. These SNPs were considered as a source of variation for various agronomic and biochemical traits in cotton (Logan-Young et al. 2015).

Conclusions

The study of SNPs opens new horizons for plant biotechnologists to improve various features of a crop plant; a single SNP has the potential to alter the expression of a gene by inducing changes in its amino acid sequence. SNPs identified in coding regions of genes have gained more attention from molecular plant breeders as comparing with those found in non-coding regions. Various assays have been exploited using these markers to detect genetic variability in the genomes of field crops. Plant researchers have utilized these markers successfully in cotton and other crops for improvement and development of tolerance to biotic and abiotic stresses, fibre quality and yield in order to enhance profitability for farmers.

Abbreviations

dCAPS: Derived cleaved amplified polymorphic sequences; GBS: Genotyping by sequencing; GWAS: Genome wide association study; NGS: Next generation sequencing; PCR: Polymerase chain reaction; QTL: Quantitative trait loci; QTN: Quantitative trait nucleotides; RAD-seq: Restriction-site associated DNA sequencing; RILs: Recombinant inbred lines; SLAF-seq: Specific locus amplified fragment sequencing; SNP: Single nucleotide polymorphism

Acknowledgements

The authors are highly grateful to reviewers for critical review and also thankful to all of collaborators for giving productive contribution for preparing this review article.

Funding

Not applicable.

Availability of data and materials

Not applicable.

Authors' contributions

Majeed S and Azhar MT has collected the literature and wrote this draft, Rana IA, Atif RM, Ali Z and Hinze L have reviewed and edited this article for publication. All authors read and approved the final manuscript.

Authors' information

Not applicable.

Ethics approval and consent to participate

Not applicable.

Consent for publication

All the authors and co-authors are agreed to submit the review article in BMC Journal of Cotton Research.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Plant Breeding and Genetics, University of Agriculture, Faisalabad, Pakistan. ²Centre of Advanced Studies in Agriculture and Food Security/Center of Agricultural Biochemistry and Biotechnology, University of Agriculture, Faisalabad, Pakistan. ³Institute of Plant Breeding and Biotechnology, Muhammad Nawaz Sharif University of Agriculture, Multan, Pakistan. ⁴US Department of Agriculture, Agricultural Research Service, College Station, TX, USA.

Received: 22 February 2019 Accepted: 30 April 2019

Published online: 29 May 2019

References

- Abdelraheem A, Fang DD, Zhang J. Quantitative trait locus mapping of drought and salt tolerance in an introgressed recombinant inbred line population of upland cotton under the greenhouse and field conditions. *Euphytica*. 2018; 214(1):8. <https://doi.org/10.1007/s10681-017-2095-x>.
- Akond M, Liu S, Schoener L, et al. A SNP-based genetic linkage map of soybean using the SoySNP6K Illumina Infinium BeadChip genotyping array. *Plant Genet Genomics Biotech*. 2013;1(3):80–9. <https://doi.org/10.5147/jpgs.2013.0090>.
- Akter T, Islam AKMA, Rasul MG, et al. Evaluation of genetic diversity in short duration cotton (*Gossypium hirsutum* L.). *J Cotton Res*. 2019;2:1. <https://doi.org/10.1186/s42397-018-0018-6>.
- Ali I, Teng Z, Bai Y, et al. A high density SLAF-SNP genetic map and QTL detection for fibre quality traits in *Gossypium hirsutum*. *BMC Genomics*. 2018; 19(1):879. <https://doi.org/10.1186/s12864-018-5294-5>.
- Alkan C, Coe BP, Eichler EE. Genome structural variation discovery and genotyping. *Nat Rev Genet*. 2011;12(5):363–76.
- An C, Saha S, Jenkins JN, et al. Cotton (*Gossypium* spp.) *R2R3-MYB* transcription factors SNP identification, phylogenomic characterization, chromosome localization, and linkage mapping. *Theor Appl Genet*. 2008; 116(7):1015–26.
- Andrews KR, Good JM, Miller MR, et al. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat Rev Genet*. 2016;17(2):81–92. <https://doi.org/10.1038/nrg.2015.28>.
- Aranzana MJ, Kim S, Zhao K, et al. Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. *PLoS Genet*. 2005;1(5):e60.
- Arruda M, Lipka A, Brown P, et al. Comparing genomic selection and marker-assisted selection for Fusarium head blight resistance in wheat (*Triticum aestivum* L.). *Mol Breeding*. 2016;36(7):84. <https://doi.org/10.1007/s11032-016-0508-5>.
- Baytar AA, Erdogan O, Frary A, et al. Molecular diversity and identification of alleles for Verticillium wilt resistance in elite cotton (*Gossypium hirsutum* L.) germplasm. *Euphytica*. 2017;213(2):31. <https://doi.org/10.1007/s10681-016-1787-y>.
- Bednarz CW, Nichols RL. Phenological and morphological components of cotton crop maturity. *Crop Sci*. 2005;45(4):1497–503.
- Bergey CM, Pozzi L, Disotell TR, et al. A new method for genome-wide marker development and genotyping holds great promise for molecular primatology. *Int J Primatol*. 2013;34(2):303–14.
- Bradov JM, Bauer PJ, Hinojosa O, et al. Quantitation of cotton fibre-quality variations arising from boll and plant growth environments. *Eur J Agron*. 1997;6(3–4):191–204.
- Buckler ES, Holland JB, Bradbury PJ, et al. The genetic architecture of maize flowering time. *Sci*. 2009;325(5941):714–8.
- Byers RL, Harker DB, Yourstone SM, et al. Development and mapping of SNP assays in allotetraploid cotton. *Theor Appl Genet*. 2012;124(7):1201–14. <https://doi.org/10.1007/s00122-011-1780-8>.
- Chen W, Yao J, Chu L, et al. Genetic mapping of the nulliplex-branch gene (*gb_nb1*) in cotton using next-generation sequencing. *Theor Appl Genet*. 2015; 128(3):539–47. <https://doi.org/10.1007/s00122-014-2452-2>.
- Chung WH, Jeong N, Kim J, et al. Population structure and domestication revealed by high-depth resequencing of Korean cultivated and wild soybean genomes. *DNA Res*. 2013;21(2):153–67. <https://doi.org/10.1093/dnares/dst047>.
- Cober ER, Molnar SJ, Charette M, et al. A new locus for early maturity in soybean. *Crop Sci*. 2010;50(2):524–7.
- Davey JW, Hohenlohe PA, Etter PD, et al. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet*. 2011; 12(7):499–510. <https://doi.org/10.1038/nrg3012>.
- De Souza GA, Arntzen MØ, Wiker HG. MSMSpddb: providing protein databases of closely related organisms to improve proteomic characterization of prokaryotic microbes. *Bioinformatics*. 2010;26(5):698–9. <https://doi.org/10.1093/bioinformatics/btq004>.
- DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011; 43(5):491–8. <https://doi.org/10.1038/ng.806>.
- Diouf L, Magwanga RO, Gong W, et al. QTL mapping of fiber quality and yield-related traits in an intra-specific upland cotton using genotype by sequencing (GBS). *Int J Mol Sci*. 2018;19(2):441. <https://doi.org/10.3390/ijms19020441>.
- Diouf L, Pan Z, He SP, et al. High-density linkage map construction and mapping of salt-tolerant QTLs at seedling stage in upland cotton using genotyping by sequencing (GBS). *Int J Mol Sci*. 2017;18(12):2622. <https://doi.org/10.3390/ijms18122622>.
- Du X, Huang G, He S, et al. Resequencing of 243 diploid cotton accessions based on an updated a genome identifies the genetic basis of key agronomic traits. *Nat Genet*. 2018;50(6):796–802. <https://doi.org/10.1038/s41588-018-0116-x>.
- Du X, Liu G, Fu H, et al. Identification and transferring breeding of nulliplex-branch germplasmes in upland cotton. *China Cotton*. 1996; 23(9):7–8.
- Elshire RJ, Glaubitz JC, Sun Q, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*. 2011; 6(5):e19379.
- Fang DD, Xiao J, Canci PC, et al. A new SNP haplotype associated with blue disease resistance gene in cotton (*Gossypium hirsutum* L.). *Theor Appl Genet*. 2010;120(5):943–53. <https://doi.org/10.1007/s00122-009-1223-y>.
- Fang L, Wang Q, Hu Y, et al. Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. *Nat Genet*. 2017;49(7):1089–98. <https://doi.org/10.1038/ng.3887>.
- Ganal MW, Altmann T, Röder MS. SNP identification in crop plants. *Curr Opin Plant Biol*. 2009;12(2):211–7.
- Gore MA, Fang DD, Poland JA, et al. Linkage map construction and quantitative trait locus analysis of agronomic and fiber quality traits in cotton. *Plant Genome*. 2014;7(1):1–10. <https://doi.org/10.3835/plantgenome2013.07.0023>.
- Handi SS, Katageri IS, Adiger S, et al. Association mapping for seed cotton yield, yield components and fibre quality traits in upland cotton (*Gossypium hirsutum* L.) genotypes. *Plant Breed*. 2017;136(6):958–68. <https://doi.org/10.1111/pbr.12536>.
- Hinze LL, Hulse-Kemp AM, Wilson IW, et al. Diversity analysis of cotton (*Gossypium hirsutum* L.) germplasm using the CottonSNP63K Array. *BMC Plant Biol*. 2017;17(1):37. <https://doi.org/10.1186/s12870-017-0981-y>.
- Hsu CY, An C, Saha S, et al. Molecular and SNP characterization of two genome specific transcription factor genes *GhMyb8* and *GhMyb10* in cotton species. *Euphytica*. 2008;159(1–2):259–73.
- Huang C, Nie X, Shen C, et al. Population structure and genetic basis of the agronomic traits of upland cotton in China revealed by a genome-wide association study using high-density SNPs. *Plant Biotechnol J*. 2017;15(11):1374–86.
- Hulse-Kemp AM, Lemm J, Plieske J, et al. Development of a 63K SNP array for cotton and high-density mapping of intra and inter-specific populations of *Gossypium* spp. G3: Genes Genomes Genetics. 2015;5(6):1187–209. <https://doi.org/10.1534/g3.115.018416>.
- Islam MS, Thyssen GN, Jenkins JN, et al. Detection, validation, and application of genotyping-by-sequencing based single nucleotide polymorphisms in upland cotton. *Plant Genome*. 2015;8(1):1–10. <https://doi.org/10.3835/plantgenome2014.07.0034>.
- Islam MS, Thyssen GN, Jenkins JN, et al. A MAGIC population-based genome-wide association study reveals functional association of *GhRBB1_A07* gene with superior fiber quality in cotton. *BMC Genomics*. 2016a;17(1):903. <https://doi.org/10.1186/s12864-016-3249-2>.
- Islam MS, Zeng L, Thyssen GN, et al. Mapping by sequencing in cotton (*Gossypium hirsutum*) line *MD52ne* identified candidate genes for fiber strength and its related quality attributes. *Theor Appl Genet*. 2016b;129(6):1071–86.
- Jang SJ, Sato M, Sato K, et al. A single-nucleotide polymorphism in an endo-1, 4- β -glucanase gene controls seed coat permeability in soybean. *PLoS One*. 2015;10(6):e0128527. <https://doi.org/10.1371/journal.pone.0128527>.
- Jia X, Pang C, Wei H, et al. High-density linkage map construction and QTL analysis for earliness-related traits in *Gossypium hirsutum* L. *BMC Genomics*. 2016;17(1):909. <https://doi.org/10.1186/s12864-016-3269-y>.
- Yu JZ, Kohel RJ, Fang DD, et al. A high-density simple sequence repeat and single nucleotide polymorphism genetic map of the tetraploid cotton genome. G3: Genes Genomes Genetics. 2012;2(1):43–58. <https://doi.org/10.1534/g3.111.001552>.
- Johnson DS, Mortazavi A, Myers RM, et al. Genome-wide mapping of in vivo protein-DNA interactions. *Sci*. 2007;316(5830):1497–502.
- Johnson J, MacDonald S, Meyer L, et al. The world and United States cotton outlook. In: *Agricultural Outlook Forum 2018*. Arlington: United States Department of Agriculture; 2018. <https://www.usda.gov/oce/forum/2018/commodities/Cotton.pdf>.
- Keerio AA, Shen C, Nie Y, et al. QTL mapping for fiber quality and yield traits based on introgression lines derived from *Gossypium hirsutum* \times *G. tomentosum*. *Int J Mol Sci*. 2018;19(1):243. <https://doi.org/10.3390/ijms19010243>.
- Kim C, Guo H, Kong W, et al. Application of genotyping by sequencing technology to a variety of crop breeding programs. *Plant Sci*. 2016;242:14–22.

- Kottapalli P, Ulloa M, Kottapalli KR, et al. SNP marker discovery in Pima cotton (*Gossypium barbadense* L.) leaf transcriptomes. *Genomics Insights*. 2016;9(GE1.S40377):51–60. <https://doi.org/10.4137/GE1.S40377>.
- Kumar NM, Katageri IS, Gowda SA, et al. 63K SNP chip based linkage mapping and QTL analysis for fibre quality and yield component traits in *Gossypium barbadense* L. cotton. *Euphytica*. 2019;215(1):6. <https://doi.org/10.1007/s10681-018-2326-9>.
- Lateef DD. DNA marker technologies in plants and applications for crop improvements. *J Biosci Med*. 2015;3(5):7–18. <https://doi.org/10.4236/jbm.2015.35002>.
- Li C, Dong Y, Zhao T, et al. Genome-wide SNP linkage mapping and QTL analysis for fiber quality and yield traits in the upland cotton recombinant inbred lines population. *Front Plant Sci*. 2016;7:1356. <https://doi.org/10.3389/fpls.2016.01356>.
- Li C, Fu Y, Sun R, et al. Single-locus and multi-locus genome-wide association studies in the genetic dissection of fiber quality traits in upland cotton (*Gossypium hirsutum* L.). *Front Plant Sci*. 2018a;9:1083. <https://doi.org/10.3389/fpls.2018.01083>.
- Li C, Wang Y, Ai N, et al. A genome-wide association study of early-maturation traits in upland cotton based on the CottonSNP80K array. *J Integr Plant Biol*. 2018b;60(10):970–85. <https://doi.org/10.1111/jipb.12673>.
- Li F, Kitashiba H, Inaba K, et al. A *Brassica rapa* linkage map of EST-based SNP markers for identification of candidate genes controlling flowering time and leaf morphological traits. *DNA Res*. 2009;16(6):311–23.
- Li L, Zhao S, Su J, et al. High-density genetic linkage map construction by F₂ populations and QTL analysis of early-maturity traits in upland cotton (*Gossypium hirsutum* L.). *PLoS One*. 2017a;12(8):e0182918. <https://doi.org/10.1371/journal.pone.0182918>.
- Li T, Ma X, Li N, et al. Genome-wide association study discovered candidate genes of Verticillium wilt resistance in upland cotton (*Gossypium hirsutum* L.). *Plant Biotech J*. 2017b;15(12):1520–32. <https://doi.org/10.1111/pbi.12734>.
- Li X, Wu M, Liu G, et al. Identification of candidate genes for fiber length quantitative trait loci through RNA-Seq and linkage and physical mapping in cotton. *BMC Genomics*. 2017c;18(1):427. <https://doi.org/10.1186/s12864-017-3812-5>.
- Liu R, Gong J, Xiao X, et al. GWAS analysis and QTL identification of fiber quality traits and yield components in upland cotton using enriched high-density SNP markers. *Front Plant Sci*. 2018;9(1067). <https://doi.org/10.3389/fpls.2018.01067>.
- Logan-Young CJ, Yu JZ, Verma SK, et al. SNP discovery in complex allotetraploid genomes (*Gossypium* spp., Malvaceae) using genotyping by sequencing. *Appl Plant Sci*. 2015;3(3):1400077. <https://doi.org/10.3732/apps.1400077>.
- Lu F, Lipka AE, Glaubitz J, et al. Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLoS Genet*. 2013;9(1):e1003215. <https://doi.org/10.1371/journal.pgen.1003215>.
- Ma JQ, Huang L, Ma CL, et al. Large-scale SNP discovery and genotyping for constructing a high-density genetic map of tea plant using specific-locus amplified fragment sequencing (SLAF-seq). *PLoS One*. 2015;10(6):e0128798.
- Ma Q, Wu M, Pei W, et al. RNA-seq-mediated transcriptome analysis of a fiberless mutant cotton and its possible origin based on SNP markers. *PLoS One*. 2016;11(3):e0151994. <https://doi.org/10.1371/journal.pone.0151994>.
- Ma Z, He S, Wang X, et al. Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. *Nat Genet*. 2018;50(6):803–13. <https://doi.org/10.1038/s41588-018-0119-7>.
- Magwanga RO, Lu P, Kirungu JN, et al. GBS mapping and analysis of genes conserved between *Gossypium tomentosum* and *Gossypium hirsutum* cotton cultivars that respond to drought stress at the seedling stage of the BC₂F₂ generation. *Int J Mol Sci*. 2018;19(6):1614. <https://doi.org/10.3390/ijms19061614>.
- Moen T, Hayes B, Nilsen F, et al. Identification and characterisation of novel SNP markers in Atlantic cod: evidence for directional selection. *BMC Genet*. 2008; 9(1):18. <https://doi.org/10.1186/1471-2156-9-18>.
- Morin PA, Luikart G, Wayne RK, et al. SNPs in ecology, evolution and conservation. *Trends Ecol Evol*. 2004;19(4):208–16.
- Palanga KK, Jamshed M, Rashid M, et al. Quantitative trait locus mapping for Verticillium wilt resistance in an upland cotton recombinant inbred line using SNP-based high density genetic map. *Front Plant Sci*. 2017;8:382. <https://doi.org/10.3389/fpls.2017.00382>.
- Pasam RK, Sharma R, Malosetti M, et al. Genome-wide association studies for agronomical traits in a world wide spring barley collection. *BMC Plant Biol*. 2012;12(1):16. <https://doi.org/10.1186/1471-2229-12-16>.
- Poland JA, Brown PJ, Sorrells ME, et al. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS One*. 2012;7(2):e32253. <https://doi.org/10.1371/journal.pone.0032253>.
- Poland JA, Rife TW. Genotyping-by-sequencing for plant breeding and genetics. *Plant Genome*. 2012;5(3):92–102. <https://doi.org/10.3835/plantgenome2012.05.0005>.
- Qi H, Wang N, Qiao W, et al. Construction of a high-density genetic map using genotyping by sequencing (GBS) for quantitative trait loci (QTL) analysis of three plant morphological traits in upland cotton (*Gossypium hirsutum* L.). *Euphytica*. 2017;213(4):83. <https://doi.org/10.1007/s10681-017-1867-7>.
- Rafalski A. Applications of single nucleotide polymorphisms in crop genetics. *Curr Opin Plant Biol*. 2002;5(2):94–100.
- Reddy UK, Nimmakayala P, Abburi VL, et al. Genome-wide divergence, haplotype distribution and population demographic histories for *Gossypium hirsutum* and *Gossypium barbadense* as revealed by genome-anchored SNPs. *Sci Rep*. 2017;7:41285. <https://doi.org/10.1038/srep41285>.
- Reitzel A, Herrera S, Layden M, et al. Going where traditional markers have not gone before: utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. *Mol Ecol*. 2013; 22(11):2953–70. <https://doi.org/10.1111/mec.12228>.
- Sekmen AH, Ozgur R, Uzilday B, et al. Reactive oxygen species scavenging capacities of cotton (*Gossypium hirsutum*) cultivars under combined drought and heat induced oxidative stress. *Environ Exp Bot*. 2014;99:141–9. <https://doi.org/10.1016/j.envexpbot.2013.11.010>.
- Semagn K, Bjørnstad Å, Ndjiondjop M. Principles, requirements and prospects of genetic mapping in plants. *Afr J Biotechnol*. 2006;5(25):2569–87.
- Shaheen T, Asif M, Zafar Y. Single nucleotide polymorphism analysis of *MT-SHSP* gene of *Gossypium arboreum* and its relationship with other diploid cotton genomes, *G. hirsutum* and *Arabidopsis thaliana*. *Pakistan J Bot*. 2009;41(1):177–83.
- Shaheen T, Zafar Y, Rahman M. Phylogenetic analysis of cotton species (diploid genomes) using single nucleotide polymorphisms (SNPs) markers. *Pakistan J Agri Sci*. 2016;53(2):283–90. <https://doi.org/10.21162/PAKJAS/16.2300>.
- Shi Z, Liu S, Noe J, et al. SNP identification and marker assay development for high-throughput selection of soybean cyst nematode resistance. *BMC Genomics*. 2015;16(1):314. <https://doi.org/10.1186/s12864-015-1531-3>.
- Silva T, Corrêa R, Castilho Y, et al. Widespread distribution and a new recombinant species of Brazilian virus associated with cotton blue disease. *Virology*. 2008;51(1):123. <https://doi.org/10.1016/j.virus.2008.05.012>.
- Song L, Koga Y, Ecker JR. Profiling of transcription factor binding events by chromatin immunoprecipitation sequencing (ChIP-seq). *Curr Protoc Plant Biol*. 2016;1(2):293–306. <https://doi.org/10.1002/cppb.20014>.
- Song X, Zhang T. Quantitative trait loci controlling plant architectural traits in cotton. *Plant Sci*. 2009;177(4):317–23.
- Su J, Fan S, Li L, et al. Detection of favorable QTL alleles and candidate genes for lint percentage by GWAS in Chinese upland cotton. *Front Plant Sci*. 2016a;7: 1576. <https://doi.org/10.3389/fpls.2016.01576>.
- Su J, Li L, Pang C, et al. Two genomic regions associated with fiber quality traits in Chinese upland cotton under apparent breeding selection. *Sci Rep*. 2016b; 6:38496. <https://doi.org/10.1038/srep38496>.
- Su J, Li L, Zhang C, et al. Genome-wide association study identified genetic variations and candidate genes for plant architecture component traits in Chinese upland cotton. *Theor Appl Genet*. 2018;131(6):1299–314.
- Su J, Pang C, Wei H, et al. Identification of favorable SNP alleles and candidate genes for traits related to early maturity via GWAS in upland cotton. *BMC Genomics*. 2016c;17(1):687. <https://doi.org/10.1186/s12864-016-2875-z>.
- Suchan T, Pitteloud C, Gerasimova NS, et al. Hybridization capture using RAD probes (hyRAD), a new tool for performing genomic analyses on collection specimens. *PLoS One*. 2016;11(3):e0151651. <https://doi.org/10.1371/journal.pone.0151651>.
- Sun X, Liu D, Zhang X, et al. SLAF-seq: an efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. *PLoS One*. 2013;8(3):e58700. <https://doi.org/10.1371/journal.pone.0058700>.
- Tan Z, Zhang Z, Sun X, et al. Genetic map construction and fiber quality QTL mapping using the CottonSNP80K array in upland cotton. *Front Plant Sci*. 2018;9:225. <https://doi.org/10.3389/fpls.2018.00225>.
- Tang B, Jenkins J, Watson C, et al. Evaluation of genetic variances, heritabilities, and correlations for yield and fiber traits among cotton F₂ hybrid populations. *Euphytica*. 1996;91(3):315–22.
- Wang H, Huang C, Guo H, et al. QTL mapping for fiber and yield traits in upland cotton under multiple environments. *PLoS One*. 2015a;10(6):e0130742.
- Wang H, Jin X, Zhang B, et al. Enrichment of an intraspecific genetic map of upland cotton by developing markers using parental RAD sequencing. *DNA Res*. 2015b;22(2):147–60. <https://doi.org/10.1093/dnares/dsu047>.

- Wang S, Chen J, Zhang W, et al. Sequence-based ultra-dense genetic and physical maps reveal structural variations of allopolyploid cotton genomes. *BMC Genome Biol.* 2015;16(1):108. <https://doi.org/10.1186/s13059-015-0678-1>.
- Wang X, Lu X, Wang J, et al. Mining and analysis of SNP in response to salinity stress in upland cotton (*Gossypium hirsutum* L.). *PLoS One.* 2016;11(6):e0158142. <https://doi.org/10.1371/journal.pone.0158142>.
- Wen T, Dai B, Wang T, et al. Genetic variations in plant architecture traits in cotton (*Gossypium hirsutum*) revealed by a genome-wide association study. *Crop J.* 2019;7(2):209–16. <https://doi.org/10.1016/j.cj.2018.12.004>.
- Xiao J, Fang DD, Bhatti M, et al. A SNP haplotype associated with a gene resistant to *Xanthomonas axonopodis* pv. *Malvacearum* in upland cotton (*Gossypium hirsutum* L.). *Mol Breeding.* 2010;25(4):593–602.
- Xie W, Feng Q, Yu H, et al. Parent-independent genotyping for constructing an ultrahigh-density linkage map based on population sequencing. *Proc Natl Acad Sci.* 2010;107(23):10578–83.
- Yano K, Yamamoto E, Aya K, et al. Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nat Genet.* 2016;48(8):927–34. <https://doi.org/10.1038/ng.3596>.
- Yu J, Buckler ES. Genetic association mapping and genome organization of maize. *Curr Opin Biotechnol.* 2006;17(2):155–60.
- Zeng YD, Sun JL, Bu SH, et al. EcoTILLING revealed SNPs in *GhSus* genes that are associated with fiber and seed-related traits in upland cotton. *Sci Rep.* 2016; 6:29250. <https://doi.org/10.1038/srep29250>.
- Zhang T, Jin Y, Zhao JH, et al. Host-induced gene silencing of the target gene in fungal cells confers effective resistance to the cotton wilt disease pathogen *Verticillium dahliae*. *Mol Plant.* 2016a;9(6):939–42. <https://doi.org/10.1016/j.molp.2016.02.008>.
- Zhang YC, Feng CH, Bie S, et al. Analysis of short fruiting branch gene and marker-assisted selection with SNP linked to its trait in upland cotton. *J Cotton Res.* 2018a;1(1):5. <https://doi.org/10.1186/s42397-018-0001-2>.
- Zhang YP, Wang QL, Zuo DY, et al. Map-based cloning of a recessive gene v_1 for virescent leaf expression in cotton (*Gossypium* spp.). *J Cotton Res.* 2018b;1(1): 10. <https://doi.org/10.1186/s42397-018-0009-7>.
- Zhang Z, Ge Q, Liu A, et al. Construction of a high-density genetic map and its application to QTL identification for fiber strength in upland cotton. *Crop Sci.* 2017;57(2):774–88. <https://doi.org/10.2135/cropsci2016.06.0544>.
- Zhang Z, Li J, Muhammad J, et al. High resolution consensus mapping of quantitative trait loci for fiber strength, length and micronaire on chromosome 25 of the upland cotton (*Gossypium hirsutum* L.). *PLoS One.* 2015;10(8):e0135430. <https://doi.org/10.1371/journal.pone.0135430>.
- Zhang Z, Shang H, Shi Y, et al. Construction of a high-density genetic map by specific locus amplified fragment sequencing (SLAF-seq) and its application to quantitative trait loci (QTL) analysis for boll weight in upland cotton (*Gossypium hirsutum*). *BMC Plant Biol.* 2016b;16(1):79. <https://doi.org/10.1186/s12870-016-0741-4>.
- Zhao Y, Wang H, Chen W, et al. Regional association analysis-based fine mapping of three clustered QTL for *Verticillium* wilt resistance in cotton (*G. hirsutum* L.). *BMC Genomics.* 2017;18(1):661. <https://doi.org/10.1186/s12864-017-4074-y>.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

