Chapter

Prospects for Molecular Breeding in Cotton, *Gossypium* spp

Ishwarappa S. Katageri, S. Anjan Gowda, Prashanth B.N, Mahesh Biradar, Rajeev M and Rajesh S. Patil

Abstract

Conventional breeding interventions in cotton have been successful and these techniques have doubled the productivity of cotton, but it took around 40 years. One of the techniques of molecular biology i.e., genetic engineering has brought significant improvement in productivity within the year of introduction. With cotton genomics maturing, many reference genomes and related genomic resources have been developed. Newer wild species have been discovered and many countries are conserving genetic resources within and between species. This valuable germplasm can be exchanged among countries for increasing cotton productivity. As many as 249 Mapping and Association studies have been carried out and many QTLs have been discovered and it is high time for researchers to get into fine-mapping studies. Techniques of genomic selection hold valuable trust for deciphering quantitative traits like fiber quality and productivity since they take in to account all minor QTLs. There are just two studies involving genomic selection in cotton, underlining its huge prospects in cotton research. Genome editing and transformation techniques have been widely used in cotton with as many as 65 events being developed across various characters, and eight studies carried out using crisper technology. These promising technologies have huge prospects for cotton production sustainability.

Keywords: cotton, wild species, reference genomes, markers, QTL mapping, genomic selection, genetic transformation, gene-editing

1. Introduction

Cotton is one among many fiber-producing species, but it is the only major crop cultivated for quenching one of the basic human necessities i.e., clothing. The ancient Harappan civilizations that were discovered in the Indus valley suggested that the first use of cotton was around the 2nd millennium BC [1]. However, the discoveries of cotton fabric at Duweilah in Jordan indicate that cotton was used as early as 4th millennium BC, but the latest discoveries at Mehrgarh in Pakistan suggest that cotton fibers were used as early as 6th millennium BC [2]. Hundreds of years ago cotton was a chief source of clothing and in the future, it would continue to be, because of its unique unparalleled qualities such as comfort, safety and eco-friendly attributes. However, with the revolution in the textile industry, the synthetic fibers were dumped into the markets with the big tag line as "cost-effective" as these synthetic fibers can be manufactured at will with the desired

fiber properties to meet the spinning demands. Synthetic fibers were assumed as a major threat to cotton cultivation but sooner than later when people realized the unsustainability, unsafe and less eco-friendly characters of the Petro-chemical based synthetic fibers, cotton is still the most preferred and produced fiber for clothing [3]. It's a big surprise that the majority of us would strictly prefer cotton-based clothing for newborn children but not synthetic fibers, which describes its safety and comfort. Nowadays, the reinforcement of natural fiber by the synthetic fibers has proved excellent in terms of improved properties of the new fiber synthesized [4]. Apart from the primary application as clothing, coarse cotton is widely used in hospitals as cotton swabs. Cotton linters (fiber <3.5 mm) are used in the paper industry along with other pulps to manufacture technical papers, art papers etc. Cotton oil extracted from seed is used in the cosmetics and paint industry, oil can also be used for consumption/cooking if the gossypol content is very low. Cottonseed meal is used as dairy feed. Apart from being economically important, the cotton fiber serves as a powerful single-celled model in studying cell wall and cellulose research. Around 100 million people are involved in cotton production with over 250 million deriving employment in transportation, ginning process, and several million people in textile manufacturing, agriculture inputs sector and cottonseed crushing, among others [5]. The cotton export value during 2017 was around 15.62 billion US$ (Rice export value; 24.99 billion US$, Wheat export value: 45.13 billion US$) [6]. Cotton has around 30% share in world textile fibers [3]. The global textile and apparel market were to the tune of 1.7 trillion US$ [7], indicating that cotton is a very important crop globally. The world population is booming and it is expected to be 8–10 billion people by 2050 [8] On the contrary, cotton productivity has also seen a rise with few but impactful breeding and molecular breeding interventions such as the introduction of hirsutum, early maturing types, introduction of intra and interspecific hybrids/derivatives and Genetically

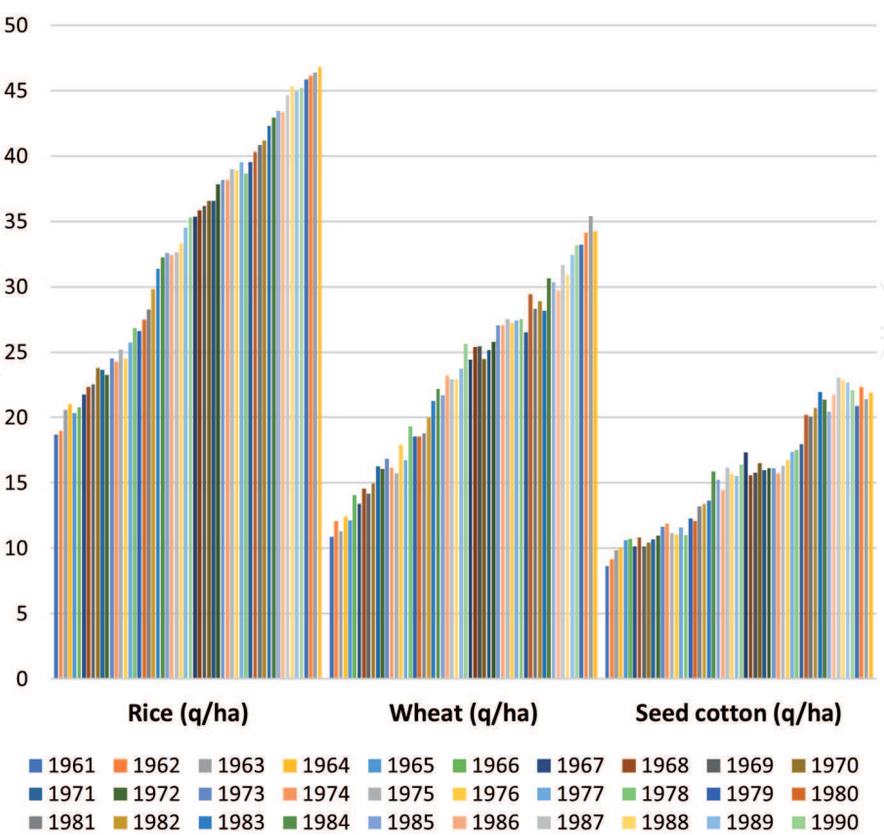

Figure 1.
Productivity improvement comparison.

Modified (GM) crops (pest and herbicide tolerance). However, the trend of increase in productivity of cotton (1961:8.62 q/ha vs. 2018:21.90 q/ha) compared to principal crops such as wheat (1961:10 q/ha vs. 2018: 34.25 q/ha) and rice (1961, 18.69 vs. 2018: 46.78 q/ha) (**Figure 1**) is very low due to less international collaboration and lesser germplasm, technology exchange. In 2050, the cotton production is required to be 94.71 MT of seed cotton (33.15 MT of lint) [9]. To meet the projected demand with the same amount of land i.e., 33 mha, there is a need to boost productivity to 28.7 t/ha of seed cotton. After the era of transgenic introductions, there is no new technological breakthrough to push the stagnant yield plateau to higher peaks. To sustain future demand with available scanty cultivable land with uncertain climatic vulnerabilities, there is a need for a strong, focused and coordinated cotton research among the world cotton research community. There lie huge prospects for molecular breeding to break the yield stagnation. Here we attempt to review the cotton genomics research carried out till date and its ability to meet the future demands.

2. Current situation

Cotton is grown in around 105 countries and total cotton production during 2018–2019 was 71.02 million tons (**Figure 2**). India, China, United States of America, Brazil, Pakistan, Turkey, Uzbekistan, Australia, Greece and Benin are the top ten producers. India is the highest producer followed by China and United States. Australia has the highest productivity of 49.05 q/ha seed cotton followed by China (45.32 q/ha) and Brazil (39.76 q/ha) (**Figure 3**). In the last fifty years, the area is more or less the same with the productivity however showing an upward trend (**Figure 2**). The percent productivity improvement graph (**Figure 4**) shows us that there were two

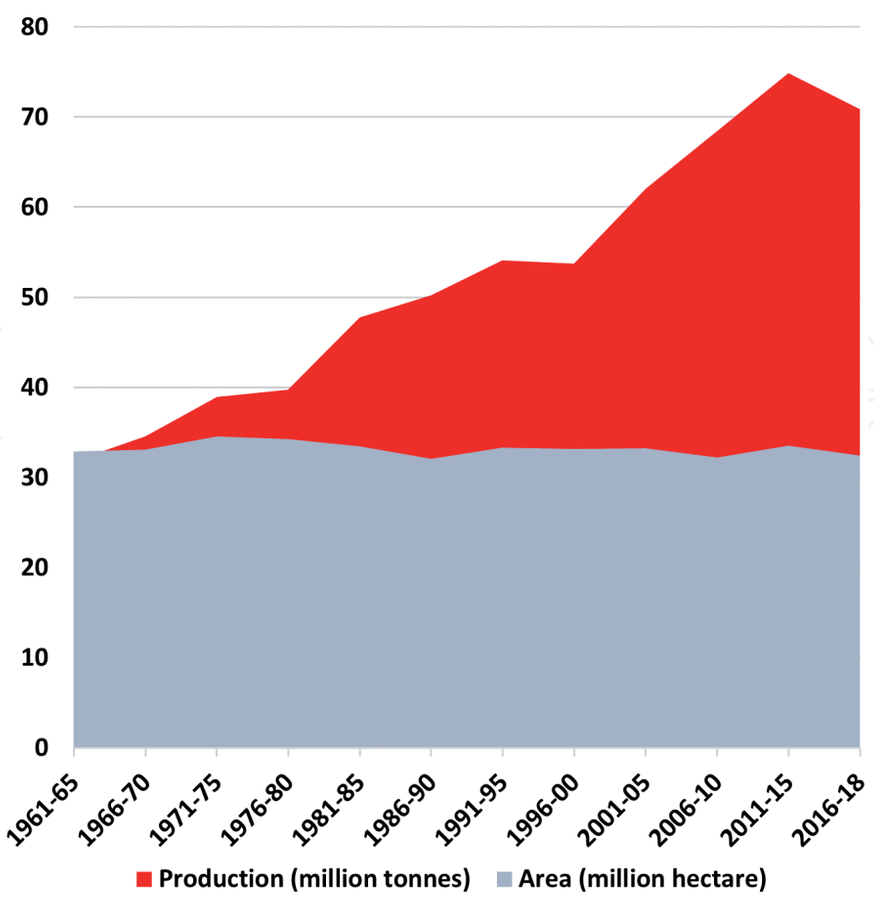

Figure 2.
World area and production statistics.

negative growth trends in the world during 1975 and 1995 but they were addressed with technological improvements. Currently, we are again witnessing a negative trend of cotton productivity growth around the world. India and USA together contribute 51% area of cotton cultivation, however the productivity (India: 13.9 t/ha & USA: 26.9 t/ha) (**Figure 5**) is low compared to other countries like Australia

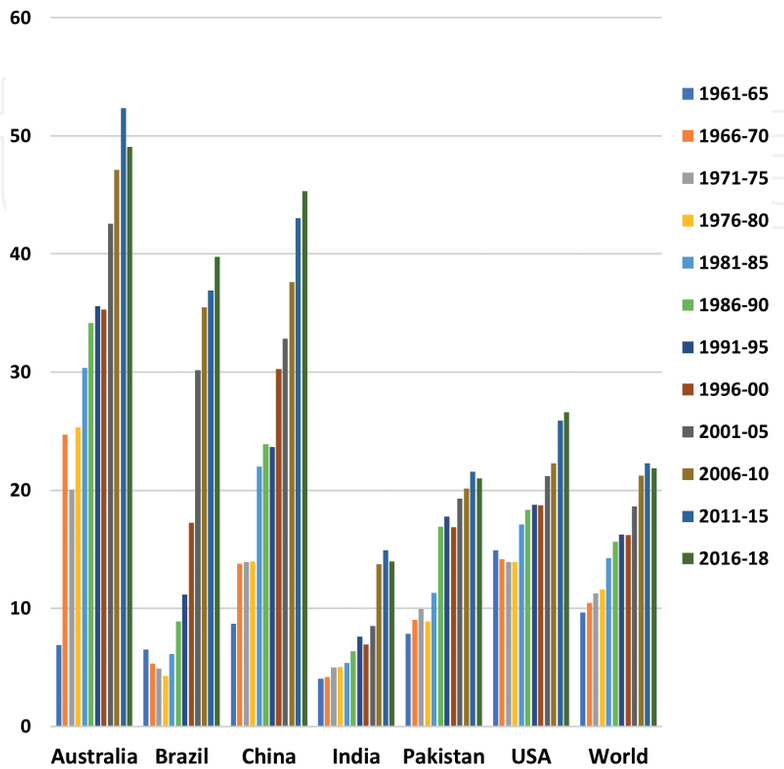

Figure 3.
World cotton productivity dynamics (q/ha).

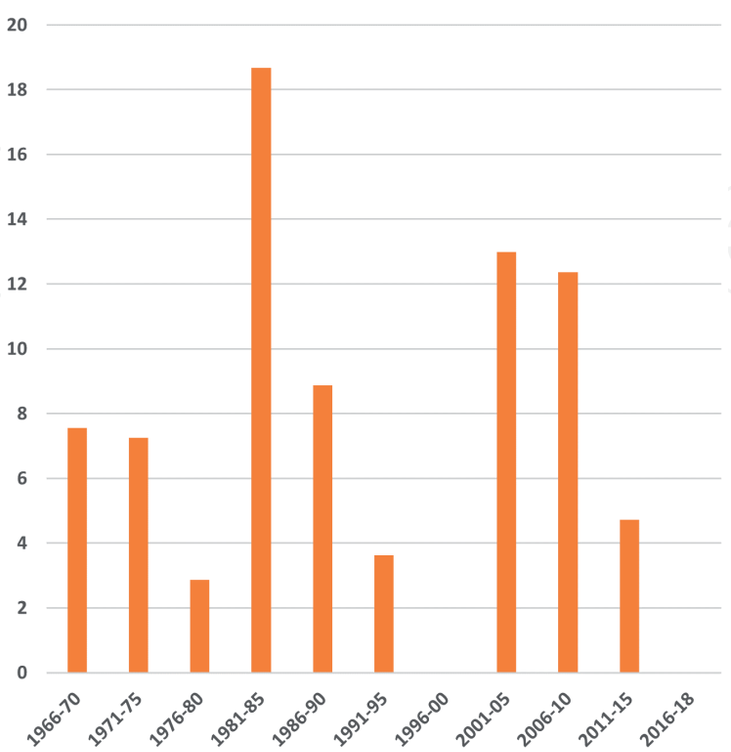

Figure 4.
Percent cotton productivity growth.

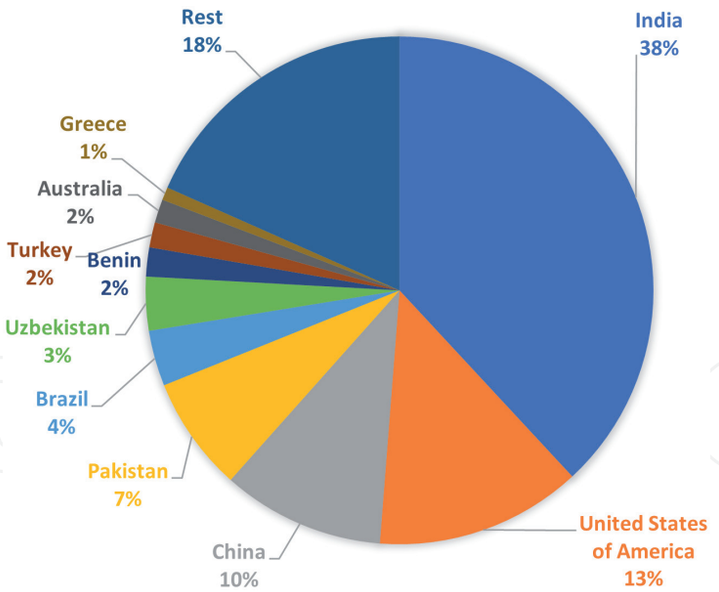

Figure 5.
World cotton growing area 2018–19 (%).

(49.05 q/ha seed cotton) and China (45.32 q/ha) [6], Improving cotton productivity in USA and India would change the outlook of cotton production sustainability. The world's two largest democracies, India & USA, have a mutual interest in promoting global security, stability and economic prosperity with cooperation at various levels. The co-operation can be extended to cotton research with other counties like Australia to make cotton productivity sustainable.

3. Genetic resource, origin, distribution and uses

Genetic variability is the main driving force of all breeding programs. The basic requirement of varietal improvement/hybrid development/marker-assisted selection is the availability of genetic resources. Even basic molecular understanding requires the occurrence of special important morphological/physiological characters upon which the studies are imposed. It is therefore necessary to understand and utilize within and between species diversity for crop improvement. The Cotton, belonging to family Malvaceae and genus Gossypium has high species diversity, which includes diploids and tetraploids with all the diploids sharing a common chromosome number i.e. 2n = 2x = 26. However, with 3-fold variation in DNA content per genome [10], they are classified into eight cytological groups (A, B, C, D, E, F, G, and H) [11–13]. All the tetraploids also share the same chromosome number i.e. 2n = 4x = 52 and have an AD genome. So, in total, the family Gossypium has around 51 recognized species, which includes 7 tetraploids and 44 diploids (**Table 1**). The genome A is thought to be originated in Asia/Africa, but the D genome is a derivative of A-genome formed by allopatric speciation i.e. due to trans-oceanic dispersion (Africa to Peru) of the A-genome. The modern-day tetraploids (AD-Genome) have originated from the trans-oceanic dispersion of A-genome to Peru followed by a polyploidization event with the native D-genome of Peru [14]. Cotton fiber is a single cell extension of the seed cell epidermis with deposition of cellulose. Only four species underwent the parallel selection pressure of domestication in America (*Gossypium hirsutum* and *G. barbadense*; tetraploids) and Africa-Asia (*G. arboreum* and *G. herbaceum*; diploids) and only these species produce the seed epidermal cell extension that is between 10 mm to 35 mm and hence are

Sl. no.	Species	Genome	Ploidy/ chromosome number	Origin	Habitat/Important traits
Primary Gene Pool					
1	*G. hirsutum*	AD1	2n = 4x = 52	Mexico	Cultivated
2	*G. barbadense*	AD2	2n = 4x = 52	South America	Cultivated, Verticillium wilt resistance
3	*G. tomentosum*	AD3	2n = 4x = 52	Hawaiian Islands	Wild, sucking pest tolerance, Drought/ Heat Resistance, Fiber strength
4	*G. mustelinum*	AD4	2n = 4x = 52	NE Brazil	Wild
5	*G. darwinii*	AD5	2n = 4x = 52	Galapagos Islands	Wild, Nematode resistance, Drought Resistance
6	*G. ekmanianum*	AD6	2n = 4x = 52	Dominican Republic	Wild
7	*G. stephensii*	AD7	2n = 4x = 52	Wake Atoll	Wild
Secondary Gene Pool					
1	*G. herbaceum*	A1	2n = 2x = 26	India	Cultivated, Drought resistance
2	*G.arboreum*	A2	2n = 2x = 26	Africa	Cultivated, Drought resistance
3	*G. anomalum*	B1	2n = 2x = 26	Africa	Wild, Fiber Length, Fiber Strength, Fiber fineness, Bollworm tolerance, Sucking pest tolerance, Bacterial Blight resistance, Drought Resistance, and Mite resistance.
4	*G. triphyllum*	B2	2n = 2x = 26	Cape Verde Islands	Wild, Jassid, Bollworm resistance, highly resistant to Bacterial blight
5	*G. trifurcatum*	B	2n = 2x = 26	Somalia	Wild
6	*G. capitis-viridis*	B3	2n = 2x = 26	Cape Verde Islands	Wild, Immune to bacterial blight
7	*G. thurberi*	D1	2n = 2x = 26	Sonora Desert	Wild, Fiber Strength, Bollworm tolerance, Fusarium wilt resistance, Frost resistance, Prolific boll bearing and high GOT.
8	*G. armourianum*	D2-1	2n = 2x = 26	Baja California	Wild, Bollworm tolerance, Sucking pest tolerance, Bacterial Blight resistance
9	*G. harknessii*	D2-2	2n = 2x = 26	Baja California	Wild, Verticillium wilt resistance, Fusarium wilt resistance, CMS male sterility source, Drought Resistance.

Sl. no.	Species	Genome	Ploidy/ chromosome number	Origin	Habitat/Important traits
10	*G. davidsonii*	D3-d	2n = 2x = 26	Baja California	Wild, sucking pest tolerance, Resistance to salinity, and Bacterial Blight resistance.
11	*G. klotzschianum*	D3-k	2n = 2x = 26	Galapagos Islands	Wild, Sucking pest resistance
12	*G. aridum*	D4	2n = 2x = 26	Pacific slopes of Mexico	Wild, arborescent, CMS male sterility source, Drought resistance, High seed index
13	*G. raimondii*	D5	2n = 2x = 26	Pacific slopes of Peru	Wild, Fiber Length, Fiber Strength, Fiber finess, Bollworm tolerance, sucking pest tolerance, Bacterial Blight resistance, Drought Resistance, High GOT.
14	*G. gossypioides*	D6	2n = 2x = 26	South Central Mexico	Wild, Resistance to Leaf Hoppers
15	*G. lobatum*	D7	2n = 2x = 26	SW Mexico	Wild, arborescent, Resistance to bollworm
16	*G. trilobum*	D8	2n = 2x = 26	West-Central Mexico	Wild, CMS male sterility source, Glabrous leaves
17	*G. laxum*	D9	2n = 2x = 26	SW Mexico	Wild, arborescent
18	*G. turneri*	D10	2n = 2x = 26	NW Mexico	Wild
19	*G. schwendimanii*	D11	2n = 2x = 26	SW Mexico	Wild, arborescent
20	*G. longicalyx*	F1	2n = 2x = 26	East Central Africa	Wild, trailing shrub, Fiber Length, Fiber fineness, and Nematode resistance
Tertiary Gene Pool					
1	*G. sturtianum*	C1	2n = 2x = 26	Central Australia	Wild, Ornamental, Fiber Strength, Fusarium wilt resistance, Cold and Frost resistance and Insensitive to photoperiod
2	*G. robinsonii*	C2	2n = 2x = 26	Western Australia	Wild
3	*G. stocksii*	E1	2n = 2x = 26	Arabian Peninsula and the Horn of Africa	Wild, Fiber Length, Fiber Strength, Drought Resistance
4	*G. somalense*	E2	2n = 2x = 26	Horn of Africa and Sudan	Wild

Sl. no.	Species	Genome	Ploidy/ chromosome number	Origin	Habitat/Important traits
5	*G. areysianum*	E3	2n = 2x = 26	Arabian Peninsula	Wild, Fiber Length, Fiber Strength, Drought Resistance
6	*G. incanum*	E4	2n = 2x = 26	Arabian Peninsula	Wild
7	*G. benadirense*	E	2n = 2x = 26	Somalia, Ethiopia, Kenya	Wild
8	*G. bricchettii*	E	2n = 2x = 26	Somalia	Wild
9	*G. vollensenii*	E	2n = 2x = 26	Somalia	Wild
10	*G. bickii*	G1	2n = 2x = 26	Central Australia	Wild
11	*G. australe*	G2	2n = 2x = 26	North Trans Australia	Wild, high GOT, Drought Resistance
12	*G. nelsonii*	G3	2n = 2x = 26	Central Australia	Wild
13	*G. costulatum*	K	2n = 2x = 26	North Kimberleys of W Australia	Wild, decumbent
14	*G. populifolium*	K	2n = 2x = 26	N Kimberleys, Australia	Wild
15	*G. cunninghamii*	K	2n = 2x = 26	The northern tip of NT, Australia	Wild
16	*G. pulchellum*	K	2n = 2x = 26	N Kimberleys, Australia	Wild
17	*G. pilosum*	K	2n = 2x = 26	NW Australia	Wild
18	*G. anapoides*	K	2n = 2x = 26	N Kimberleys, Australia	Wild

Sl. no.	Species	Genome	Ploidy/ chromosome number	Origin	Habitat/Important traits
19	*G. enthyle*	K	2n = 2x = 26	N Kimberleys, Australia	Wild
20	*G. exgiuum*	K	2n = 2x = 26	N Kimberleys, Australia	Wild, prostrate,
21	*G. londonderriense*	K	2n = 2x = 26	N Kimberleys, Australia	Wild
22	*G. marchantii*	K	2n = 2x = 26	Australia	Wild, decumbent
23	*G. nobile*	K	2n = 2x = 26	N Kimberleys, Australia	Wild
24	*G. rotundifolium*	K	2n = 2x = 26	N Kimberleys, Australia	Wild, prostrate

Table 1.
Species diversity in cotton and their importance.

cultivated for lint purpose. The rest of the species produce lint less than 10 mm with varying shades of brown to white. Some of the recent updates in number of species in the Gossypium family include *G. trifurcatum* being tentatively placed in the B genome [15]. *G. lanceolatum* was proved to be a domesticated form of *G. hirsutum* and it does not hold a species status [16]. *G. stephensii* and *G. ekmanianum* are the two new tetraploids discovered with a species status [17, 18]. Wild species are considered as the treasure of important genes required to combat biotic and abiotic stress [19]. The various species of the Gossypium genus and their important traits of interest are presented in **Table 1**.

G.arboreum and *G. herbaceum* species known as old-world cotton were majorly grown in the Indian sub-continent. The Indus valley discoveries prove that cotton was grown as early as in 6 millennium BC, with the use of cotton being mentioned in Rig Veda (15 century BC) and Manu's Dharmashastra (800 BC) [14]. From the Indian sub-continent cotton has spread to Mesopotamia, Egypt and Nubia. During the first century, cotton was introduced to Europe by the Arab traders. The East India Company that colonized India (1757) and started ruling were the biggest importers of raw cotton and they used to sell the finished goods to India and the world. *G. arboreum var neglecta* grown in Bengal was known to produce lint that could be spun to 480 counts yarn and made into Muslins which was a result of both a beautiful skill set and the cotton germplasm that were available. Garments produced here were called "webs of woven wind" [20]. However the East India Company that wanted to sell only their finished cotton garments, chopped off the thumbs of weavers and with the weavers lost, the germplasm too vanished from the world forever [21]. Though the polyploidization event of tetraploids happened in Peru, the *Gossypium hirsutum* and *G. barbadense* also called new world cottons originated in Mexico and Peru, respectively, from where it spread to both South and North America. The arrival of European colonists hastened the spread of the new world cotton to the rest of the world [22]. The East Indian Company also bought and introduced early maturing and high yield *G. hirsutum* cotton to India with many unsuccessful attempts being made (1790 in Bombay & Madras, 1840–42 in Deccan, Konkan and Hubli, 1853 in Punjab). However, the most significant development in terms of the spread of cotton was the introduction of Cambodia variety in Tamil Nadu region [23]. At present, *G. hirsutum* is cultivated in 95% cotton area due to its high yield ability. *G. barbadense* is the best source for fiber quality improvement of *G. hirsutum* as *G. barbadense* is known to produce lint

Country	*G. hirsutum*	*G. barbadense*	*G. herbaceum*	*G. arboreum*	Other species	Total	Reference
India	8851	536	565	2053	330	12,335	[25]
Uzbekistan	13,241	3019	1495	1185	31	18,971	[26]
United States of America	6302	1584	194	1729	502	10,311	[27]
China	7752	633	18	433	32	8868	[28]
Russia	4503	1057	336	365	15	6261	[29]
Brazil	1660	1509	19	219	889	4296	[29]
France	2173	483	50	69	294	3069	[29]
Australia	1573	99	39	211	22	1944	[30]

Table 2.
Country-wise list of germplasm maintained.

that can be spun to 80–120 counts yarn [24]. However, owing to their low yields, *G. arboreum*, *G. herbaceum,* and *G. barbadense* are not grown widely. Maintaining germplasm and utilizing within-species variation are big challenges in varietal development as it will be an expensive proposition. The germplasm maintained elsewhere in different countries can be efficiently utilized in breeding programs. The list of germplasm preserved is mentioned in **Table 2**. Commendable efforts are needed in pre-breeding to utilize the rare alleles/genes present in wild species. In the principal crops like rice and wheat, the IRRI and CIMMYT, respectively, are taking up the pre-breeding work and the genetic materials are being supplied to breeders around the world. Some notable works concerning cotton pre-breeding include developing populations involving genes/segments/whole chromosomes from wild species. RHMBHMTUP-C4 a random mated population was developed involving *G. hirsutum*, *G. barbadense*, *G. mustelinum* and *G. tomentosum* [31]. RMBUP-C4 was developed from crossing three elite hirsutum lines with 18 chromosome substitution lines from *G. barbadense* [32]. There is huge scope for pre-breeding work in cotton to combat biotic and abiotic stresses.

4. Cotton conventional breeding

Conventional breeding is the base for any trait improvement and without proper knowledge of conventional breeding techniques advanced molecular breeding techniques would surely lead to costly mistakes. Until the advent of molecular marker technologies, conventional breeding was the sole method for genetic improvement. Some of the popular conventional interventions in cotton were the development of determinate growth types. Development of early maturing types (resistant to boll weevils), the inclusion of morphological traits such as fergo bracts, glabrous leaves, nectariless, high gossypol for resistance to boll weevils and bollworms then followed.

Australian Conventional Breeding: American bollworm, Bacterial blight and Verticillium wilt were the major problems in Australia. Okra leaf types were used in an Australian breeding program by Norm Thomson to develop a variety called Siokra1-1 which was the first okra leaf type along with bacterial blight resistance in 1985. Dr. Peter Reid released Verticillium wilt resistance variety which was popular outside Australia [33].

Indian Conventional Breeding: Introduction of *G. hirsutum* during the 1970s and development of first intra-hirsutum hybrid cotton (H4) in India by C T Patel in 1970 [34] and Development of first interspecific hybrid (*G. hirsutum* x *G. barbadense*) in cotton (Varalaxmi) by B.H. Katarki in 1972 [35] led to utilizing hybrid vigour for higher productivity in the Indian subcontinent. Morphological traits such as fergo bracts, glabrous leaves, nectariless, and antibiosis [36–40] were included in the breeding program for bollworm tolerance. Wild species were used widely used through introgression breeding in developing novel varieties like Badnawar-1, Khandwa-1, and Khandwa-2 from *G. hirsutum* x *G. tomentosum* cross, Arogya and PKV081 using *G. hirsutum* x *G. anomalum* cross, Devitej using *G. hirsutum* x *G. herbaceum* cross, SRT-1, Deviraj and Gujarat 67 using *G. hirsutum* x *G. arboreum* cross. MCU2 and MCU5 from *G. hirsutum* x *G. barbadense* cross [41].

US Conventional breeding: Boll weevils were a major threat to cotton cultivation and elimination of boll weevils by developing early maturing short-staple types was one of the significant interventions [20]. To reduce the negative association between yield and fiber quality, exotic germplasm was used in breaking the association and varieties like MD51ne with higher fiber quality were developed [42]. Development of sub okra, smooth leaf, and nectariless for reducing tarnished plant bug populations [43] was another achievement. Some of the private companies like Delta Pine

made huge advances in cotton productivity improvement in cotton. They released mechanical harvest suitable variety called Delta pine smooth leaf which had around 25% US cotton area by 1963, Deltapine-16 with improved disease resistance and better fiber quality had around 28% area in the US by 1972, Delta pine Acala 90 premium quality cotton released was used as parental germplasm in development of many other varieties globally, all these interventions of Delta pine improved cotton production of United States significantly.

Uzbekistan Conventional Breeding: Turkestan Breeding Station established in 1992 with a major emphasis on the collection of cotton germplasm under the leadership of Dr. Zaitsev and Dr. Mauer was a major milestone in realizing the huge germplasm collection of present-day Uzbekistan. Early maturing types AK-Djura and Dehkam by Dr. Zaitsev were an important contribution. Utilizing *G. harbadense* as a resistant source to fusarium wilt and large boll types led to the release of 35-1 and 35-2 cultivars. Termez-14 high yielding cultivar developed by Dr. Ibragimov was another breakthrough. Development of Verticillium wilt resistance variety C-6524 by Dr. Alexander Avtonomov and Dr. Vadim Avtonomov had occupied more than two hundred thousand hectares for fifteen years till 2004 [44].

All these interventions along with production technologies have improved world productivity from 9.65 q/ha in 1960 to 16.20 q/ha in 2000 in the span of 40 years [6]. Conventional breeding methods are effective in transferring the traits but take considerably higher time, resources and uncertainty in the transfer of the trait. A breeding program involving large entries involves more samples needed to be tested for traits like fiber quality/oil content in cotton. We need to put a lot of resources and time to wait for the maturity of cotton. On the contrary, the Bt cotton technology, one of the spin-off technologies of genomics and molecular breeding played a significant role along with other production technologies and has achieved 5 q/ha improvement in just five years (**Figure 4**). Thus, cotton genomics/marker-assisted selection has huge potential in reducing a considerable amount of time, resource and assist conventional breeding in achieving future demands.

5. Advances in cotton genome sequencing

Having a reference genome is a boon since it is possible to characterize gene/gene families that are species-specific and which are further amenable to functional genomics work [45]. Since the publication "Toward Sequencing Cotton (Gossypium) Genomes" in 2007 by Chen and Co-workers [10], the framework was laid down for genome sequencing of cotton. The initial framework was to get first sequence of D-genome (*G. raimondii*) followed by A and then AD genome. [46] developed the first assembly in cotton (D-Genome). Now, as many as nine assemblies of *G. hirsutum*, four assemblies of *G. barbadense*, three assemblies of *G. arboreum*, three assemblies of *G. raimondii* and one each assembly of wild species such as *G. australe*, *G. darwinii*, *G. longicalyx*, *G. mustelinum*, *G. tomentosum* and *G. turneri* have been documented in CottonGen. The assembly statistics of various Gossypium genomes assembled are presented in **Table 3**. The reference genomes of cotton were used to produce a total of 17,224,361 SNPs that are documented in CottonGen by various researchers, the gene annotations and physical maps provided are valuable information for cotton scientists for various studies like the development of linkage maps, GWAS, validating the linkage maps, expression studies, development of guide RNAs in gene editing, genome-wide characterization studies of gene families etc. Since the sequencing cost is reducing day by day there are huge prospects for developing newer assemblies to catch all the variation

Reference	**[47]**	**[48]**	**[48]**	**[49]**	**[50]**	**[51]**	**[52]**	**[53]**
Species	*G. hirsutum*	*G. hirsutum*	*G. hirsutum*	*G. hirsutum*	*G. hirsutum*	*G. hirsutum*	*G. hirsutum*	*G. hirsutum*
Cultivar	TM-1	TM1	ZM24	TM-1	TM1	TM-1	TM1	TM1
Number of contigs	1235	1283	3718	4831	6733	4746	265,279	44,816
N50 of Contigs (kb)	5020	4760	1976	113.02	7839	1891(L50)	34	80
No. of scaffolds	342	599	2238	48	1025	2190	40,407	8591
N50 of scaffolds (Mb)	—	—	—	15.510	108.1	97.73 (L50)	1.6	0.764
Total assembled genome size (Mb)/Scaffold length (Mb)	2290	2286	2309	2295.26	2305.2	2347.01	2432.7	2173
Number of annotated protein coding genes	74,350	73,624	73,707	72,761	75,376	70,199	70,478	76,943

Reference	**[50]**	**[49]**	**[51]**	**[54]**	**[47]**	**[55]**	**[56]**	**[47]**
Species	*G. barbadense*	*G. barbadense*	*G. barbadense*	*G. barbadense*	*G. arboreum*	*G. arboreum*	*G. arboreum*	*G. herbaceum*
Cultivar	3–79	Hai7124	3–79	3–79	SXY1	SXY1	SXY1	Mutema
Number of Contigs	4766	6902	4930	—	2432	8223	40,381	1781
N50 of Contigs (kb)	1800	77.66	2151.56 (L50)	—	1832	1100	72	1915
No. of scaffolds	2048	29	3032	29,751	1269	—	7914	732
N50 of scaffolds (Mb)	93.8	23.44	92.88 (L50)	0.260	—	—	0.665	—
Total assembled genome size (Mb) / Scaffold length (Mb)	2195.8	2224.98	2266.65	2573.19	1637	—	1694	1556
Number of annotated protein coding genes	74,561	75,071	71,297	80,876	43,278	40,960	41,330	43,278

Reference	**[57]**	**[58]**	**[46]**	**[59]**	**[50]**	**[60]**	**[50]**	**[50]**	**[57]**
Species	*G. raimondii*	*G. raimondii*	*G. raimondii*	*G. australe*	*G. darwinii*	*G. longicalyx*	*G. mustelinum*	*G. tomentosum*	*G. turneri*
Cultivar	D5-4	D5-3 (CMD10)	—	G2-lz	AD5-32, no. 1808015.09	F1-1	1408120.09, 1408120.10, 1408121.01, 1408121.02, 1408121.03	7179.01,02,03	D10-3
Number of contigs	187	41,307	19,735	2598	821	17	2147	750	220
N50 of contigs (kb)	6291.83	44.9	135.6	1825.35	9100	95,880	2300	10,000	7909.23
No. of scaffolds	—	4715	1033	650	334	—	383	319	—
N50 of scaffolds (Mb)	58.81	2.2	6.0	143.60	101.9	—	106.8	102.9	60.46
Total assembled genome size (Mb)/Scaffold length (Mb)	734.88	775.2	761.4	1752	2183	1190.67	2315	2193.6	755.20
Number of annotated protein coding genes	40,743	40,976	37,505	40,694	78,303	38,378	74,699	78,338	38,489

Table 3.
Genome assembly statistics of various cotton species.

(within and between species) for identifying rare alleles/genes that would help us to sustain future demands in cotton improvement.

6. Transcriptome studies in cotton

The technique of isolating and characterizing the Spatio-temporal pool of mRNA to study the differential gene expression patterns between contrasting genotypes and understanding underlying alternative pathways for specific trait supremacy has been well implemented in cotton. Many characters have been targeted to find the key responsible genes and pathways, especially fiber initiation, and elongation being highly focused up on [61–68]. Other characters like Green and Brown colored cotton [69–71], Cadmium tolerance [72], Cold stress [73], Drought stress [74], Nematode resistance [75], Semigamy in Pima cotton [76], Whitefly mediated cotton leaf curl infection transcriptome [77], the transcriptome of Mepiquat chloride-induced compact types using in cotton [78] have also been done in cotton. Many studies have been carried out to identify genes that are differentially expressed, however paucity of the causes of the differential expression, hints us towards epigenetic regulations and transcription factors as the probable cause. Very few methylation studies have been carried out in cotton for fiber quality [79], male sterility [80], cold stress [81], salt tolerance [82], fruiting branch development [83]. Thus, there is huge scope for studies like differential methylation, studies on transcription factor, and correlating them with the differential expression patterns.

7. Genetic markers in cotton

In cotton, various markers like restriction fragment length polymorphism (RFLP) [84], random amplified polymorphic DNA (RAPD) [85], amplified fragment length polymorphism (AFLP) [86], simple sequence repeats microsatellites (SSRs) [87, 88], sequence-related amplified polymorphism (SRAP) [89, 90], target region amplified polymorphism (TRAP) [91], inter simple sequence repeats (ISSRs) [92], expressed sequence tag-Simple Sequence Repeat (EST-SSRs) [93] and single nucleotide polymorphism (SNP) [94–97] have been used for various genomic studies with each marker system having its advantages and disadvantages. However, SSRs were initially thought that they were sufficiently polymorphic but with the advent of high throughput SNPs, they are being less used. In the era of sequencing, the availability of the cotton reference genome is a boon to cotton researchers as a large number of SNPs are identified using whole-genome re-sequencing and transcriptome sequencing. Further, reduction in cost and advent of reduced representation sequencing methods like Genotype by sequencing (GBS) and Specific-locus amplified fragment sequencing (SLAF) provide scope for high throughput genotyping. To date, in CottonGen, 7,870,031 SSRs and 17,224,361 SNP markers are available for researchers for various studies [98]. There are few SNP arrays developed in cotton like 63k cotton array [99], 80k SNP array [100], and 50k array by Samir Sawanth and I.S. Katageri (unpublished) which are being utilized for linkage and association mapping studies. However, the SNPs associated with various traits identified using different techniques can't be used in a minimalistic laboratory with minimal cost involved, thus it is necessary to exploit trait-associated SNPs through different marker systems like CAPS (Cleaved Amplified Polymorphic Sequences) and dCAPS (derived CAPS) which require minimal laboratory set up. CAPS and dCAPS can be used as dominant marker systems and can be carried out in simple agarose gel electrophoresis. They are highly stable as they are specifically

designed for certain genomic targets [101–103]. However, there are only a few CAPs and dCAPS markers developed in cotton. There are huge prospects for developing simple PCR-based markers in cotton so that the breeders working in a remote research station with minimal laboratory can take advantage of DNA markers in cotton.

8. Molecular mapping and quantitative trait mapping

The Quantitative Trait Loci identification helps in finding the association between a marker and measurable phenotype at the genomic level or understanding the genetics of traits under study. Various types of populations like F2 [104], Recombinant inbred lines (RILs) [105], Backcross inbred lines (BILs) [106] and Multi-parent Advanced Generation Inter Cross (MAGIC) [107] are commonly used in cotton. Bi-parental RIL Mapping is one of the most common methodologies successfully employed for identifying QTLs in cotton for various traits. Genome-wide association study is also used for developing genetic maps and developing an association between the trait and DNA markers in cotton germplasm. This technique allows detecting association among various markers and traits through assessing Linkage disequilibrium (LD-mapping). In cotton the construction of linkage maps and detection of QTLs for various economic traits has been in progress since 1994 with the first RFLP linkage map [84] being published after which many maps have been constructed [94, 96, 97, 105, 108]. Many genome-wide association studies have also been carried out [95, 107, 109]. Currently, there are around 249 QTL mapping and association studies using various populations and germplasm (**Table 4**), QTLs identified using Bi-parental mapping/GWAS are presented in **Table 5**. However, QTLs discovered for various studies indicate that Chromosomes 5, 7, 10 and 25 are harboring many QTLs for fiber length, similarly Chromosomes 7 and 21, for fiber strength. For yield (Seed cotton yield/Lint yield) Chromosomes 1, 13 and 26 seem to be very important. For

Sl. no.	Genome	Number of maps
1.	*G. hirsutum* x *G. barbadense*	57
2.	*G. barbadense* x *G. hirsutum*	9
3.	*G.hirsutum* x *G.hirsutum*	99
4.	*G. barbadense* x *G. barbadense*	4
5.	*G. hirsutum* x *G. anomalum*	1
6.	*G. trilobum* x *G. raimondii*	1
7.	*G. australe* x *G. nelsonii*	1
8.	*G. hirsutum* x *G. darwinii*	1
9.	*G. hirsutum* x *G. mustelinum*	2
10.	*G. darwinii* x *G. darwinii*	1
11.	*G. davidsonii* x *G. klotzschianum*	1
12.	*G. hirsutum* x *G. tomentosum*	4
13.	*G. thurberi* x *G. trilobum*	1
14.	MAGIC (Multi parent advanced generation intercross)	6
Total		188

Table 4.
Number of Documented Quantitative Mapping studies in cotton.

Sl. no.	Trait	Documented QTLs
1.	Fiber strength	1123
2.	Fiber length	1147
3.	Boll weight	676
4.	Boll number	186
5.	Yield	275
6.	Lint Percentage	878
Total		4285

Table 5.
Number of QTLs identified for major quality and yield traits in cotton.

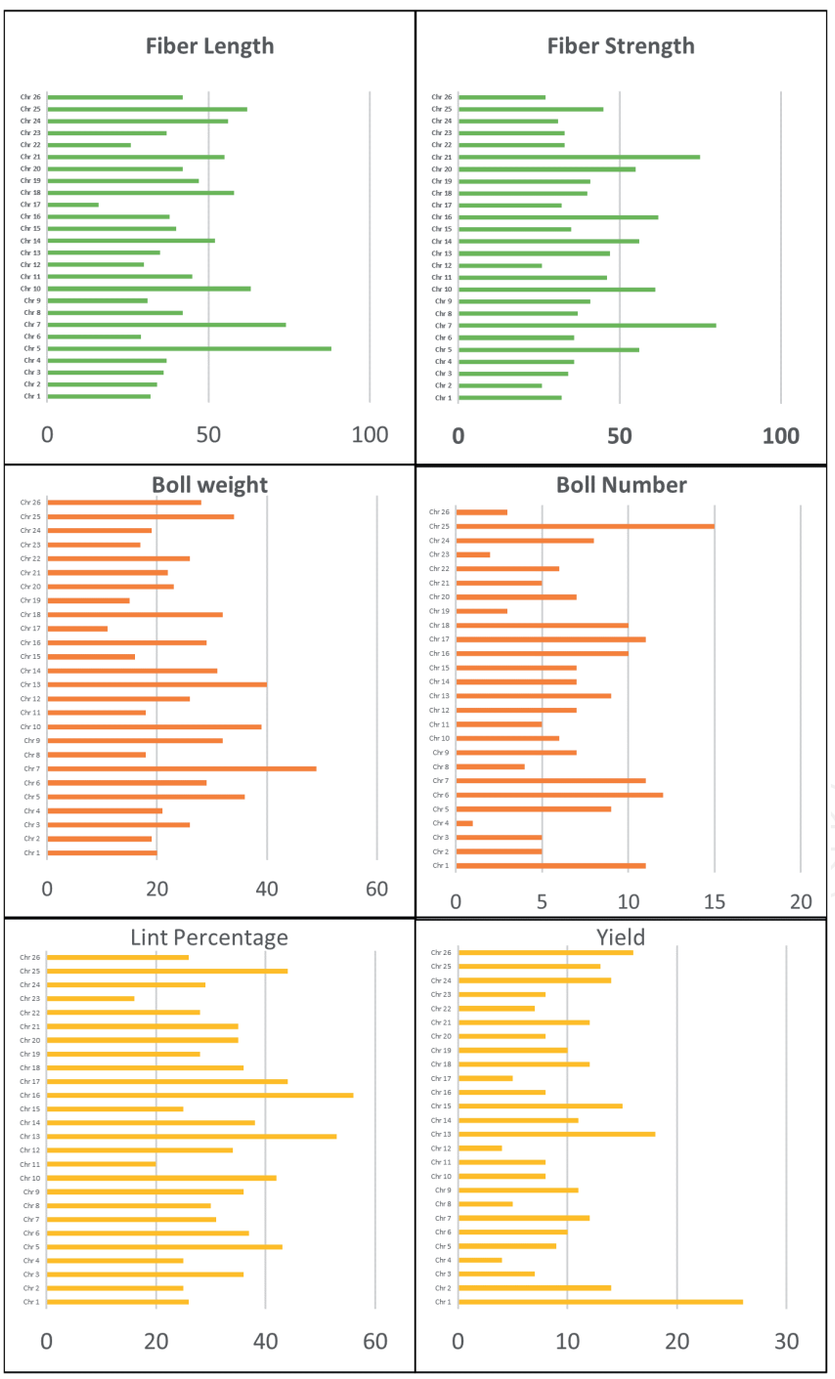

Figure 6.
QTLs identified for fiber length, strength, boll weight and yield in cotton.

boll weight, Chromosomes 7 and 13 harbor many QTLs as reported from various studies. For Boll number, Chromosome 25 and for Lint percentage Chromosomes 16 and 13 are over represented (**Figure 6**). Efforts have been made to develop linkage maps in wild species like *G. hirsutum* X *G. anomalum* [110], *G. trilobum* X *G. raimondii* [111], *G. nelsonii* x *G. austral* [112], *G. hirsutum* X *G. darwinii* [113], *G. hirsutum* X *G. mustelinum* [114], *G. darwinii* X *G. darwinii* [115], *G. klotzschianum* X *G. davidsonii* [116], *G. hirsutum* x *G. tomentosum* [117–119] and *G. thurberi* x *G. trilobum* [104]. QTLs after validation can be used directly for marker-assisted selection. Transfer of QTL/pyramiding of QTLs is one way of realizing targeted trait introgression [120] or these QTLS can be utilized for fine mapping and map-based cloning before marker-assisted selection. However, only a few validation studies are done for the Virescent gene in Virescent mutants [121, 122], the fuzzless gene in the fuzzless mutant [123], traits like fiber length [124], Fiber strength [125], leaf shape [126] and QTL affecting root-knot nematode multiplication [127] etc. Though fine-mapping is done it would require still more concentrated efforts to dissect out the traits. There are no successful cotton cultivars deployed in the field that are developed using the identified QTLs unlike in crops like Rice (MAS 946-1, Swarna Sub-1 and Cadet) and Wheat (Patwin, Expresso and AGS2026). Now that the marker development and QTL mapping has been done to a greater depth in cotton, at least for major traits like fiber quality and yield, the focus around the world should now be on utilizing all the major QTLs identified in fine mapping and then in marker assisted selection.

9. Genomic selection in cotton

QTL mapping and genome-wide association studies have identified many genomic regions responsible for the important agronomic and fiber quality traits in cotton. Among them only a few traits like disease resistance and pest resistance were qualitatively governed by a few genes/QTLs with a major effect. Marker-assisted selection (MAS) is well-suited for handling these traits. But in the majority of the crops and also cotton, most of the yield, yield contributing and fiber quality traits are quantitatively governed by one or few QTLs with relatively large effects along with several QTLs with small effects, which are not captured through QTL mapping [128, 129]. Hence targeted phenotype has not been achieved successfully through Marker Assisted Selection. Under such a situation, genomic selection (GS) would seem to be a promising and powerful tool of genomics to breed for these traits. GS is a unique form of MAS, here the basis of selection is the genotypic data on marker alleles covering the entire genome, irrespective of whether the effects associated with these marker loci are significant or not [130]. Based on these marker effect estimates, genomic estimated breeding values (GEBVs) of different individuals/lines will be calculated without actually phenotyping them, which forms the basis of selection (**Figure 7**). GS empirical studies in maize (*Zea mays*; [132–135]), rice (*Oryza sativa*; [136–139]), wheat (*Triticum aestivum*; [140–144]), and sorghum (*Sorghum bicolor*; [145–147]) have all recently shown how GS has become an efficient approach in crop breeding with recent developments in the implementation of various high-density array-based DNA marker technologies and their reduced genotyping costs. There are many marker effects estimation models that have been developed for the GS. Their predictability mainly depends on factors like marker density, training population size, and the relationship between training and breeding populations [131, 148]. Hence, the model which is capable of giving the highest GEBV accuracy will be selected. To date only two cotton GS studies have been reported. Islam et al. (2020) compared prediction ability (PA) and prediction accuracy (PACC) of Several GS models in cotton including genomic BLUP (GBLUP), ridge regression BLUP (rrBLUP), BayesB,

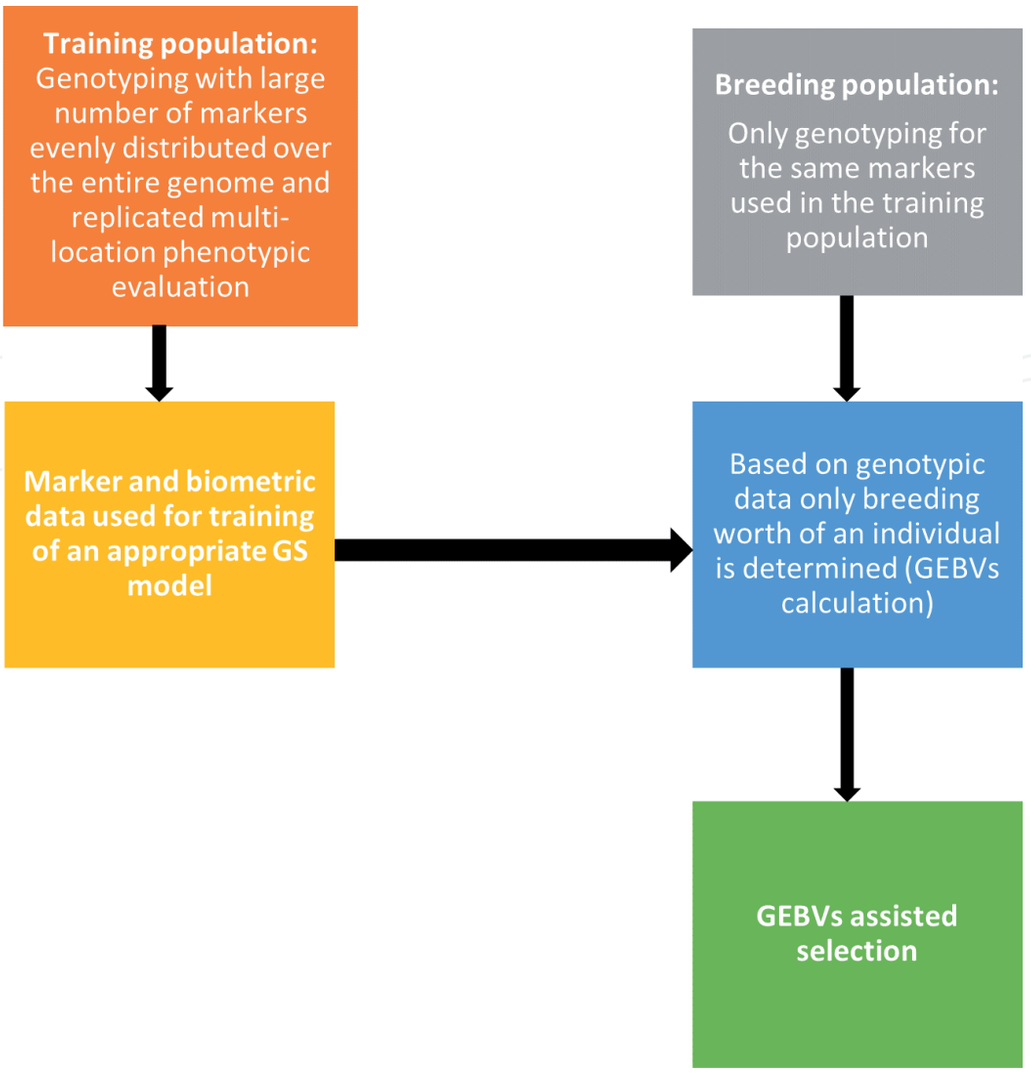

Figure 7.
Schematic representation of genomic selection (GS) scheme (based on [131]).

Bayesian LASSO, and reproducing kernel Hilbert spaces (RKHS). And reported BayesB predicted the highest accuracies among the five GS methods tested and also the same model is suggested by Gapare et al. in 2018 in cotton. In many field crops for different traits, GS prediction accuracies of >0.80 have been reached [149, 150], but now in cotton, the accuracy of 0.71 and 0.59 for fiber length and strength has been achieved, respectively [150]. The prediction ability (PA) and prediction accuracy (PACC) was 0.65, and 0.69, respectively for fiber elongation [148]. In most plant breeding programs, especially in cotton, GS is still in its infancy and one of the biggest barriers to the implementation of GS in practical plant breeding is the high start-up cost required for accurate phenotyping, maintaining a large training population and costs of genotyping entire breeding populations. However, nowadays the genotyping costs are continually decreasing and genotyping of large plant populations is much more manageable than going in for conventional phenotyping. Soon, at points in the breeding program where selection using conventional methods is too costly and time-consuming, GS may have its greatest potential usage.

10. Transgenics

With the advent of recombinant DNA technology in the 1970s, the genetic manipulation of plants entered a new age. Genes and traits previously unavailable

through traditional breeding became available through DNA recombination and with greater specificity than ever before. This modern genetic engineering technology allows the transfer of genetic material across a wide range of species and has removed the traditional limits of crossbreeding. It involves the transfer of desired genes into the plant genome, and then regeneration of a whole plant from the transformed tissue/cell. For successful development of transgenic plants, identification of suitable target tissue and efficient gene transfer protocol are essential. Therefore, understanding the genetic variability of different crop plants and genotypes for in vitro regeneration and optimization of routine regeneration protocol is prerequisite for the utilization of transformation technology in any crop. Currently, the most widely used method for transferring genes into plants is Agrobacterium-mediated transformation [151–153] and particle bombardment method [154]. Other methods, such as polyethylene glycol (PEG) mediated transformation [155] and electroporation [156] have also been used to transfer genes into plants. Cotton is a recalcitrant crop to generate from in vitro tissue cultures. Compared with many other crops, it is more difficult to obtain somatic embryogenesis, shoot multiplication and plant regeneration in cotton. The nature of tissue explants, the genetic makeup of the crop plant and presence of different growth hormones have a direct effect on regeneration potential. Genotype dependent genetic transformation is well studied and used commercially in cotton. Coker genotypes, which are amenable for regeneration in vitro by somatic embryogenesis, are widely used in genetic transformation experiments [151–153, 157]. Genotype independent genetic transformation techniques although developed [152, 158] show very low frequency of heritable gene incorporation. In the beginning, the two major goals of genetic engineering in cotton were to confer insect resistance and tolerance to more environmentally acceptable herbicides [159]. To date, 65 plus transgenic cotton events approved in India and all over the world. Continuous exposure of bollworms to BT cotton has led to resistance in them and thereby affecting the efficiency of controlling them. Cotton bollworm P450 monooxygenase gene (CYP6AE14) gene was silenced to impair larval tolerance to gossypol through the plant-mediated RNAi approach [160]. Genetic engineering is a remarkable breakthrough in modern crop improvement. Bt cotton came at the most opportune time when bollworms were causing a lot of destruction to the cotton crop making farmers helpless. Since its release in the USA in 1995, China during 1997 and in India during 2002 the Bt. technology has had a significant impact on bollworm control and reduction in usage of pesticides has been seen.

Acceptance of genetically altered cotton in various regions of the world is offering new opportunities for improvement of cotton yield and quality. Overexpression of GhUGP1 (Cotton uridine diphosphate glucose pyrophosphorylase) in upland cotton improves the fiber quality and reduces fiber sugar content [161]. Overexpression of novel sucrose synthase GhSusA1 gene leads to a considerable increase in biomass and fiber length with a moderate increase in fiber strength [162]. A silkworm fibroin gene was used to improve the fiber structure and quality [163]. The transgenic cotton plants expressing the fiber expansin gene (GhEXPA8) showed a significant improvement in fiber lengths and micronaire values [63]. The fiber quality QTL-associated phytochrome PHYA1 gene was targeted through RNAi to explore the biological roles of PHYA1 and (indirectly) other phytochrome genes in cotton [164]. The elimination of gossypol from cottonseed has been a long-standing goal of geneticists. A cotton variant was obtained using antisense technology against (+)-delta-cadinene gene to suppress terpenoid aldehydes (gossypol) but with lysigenous glands [165]. RNAi-knockdown of delta-cadinene synthase gene(s) was used to engineer plants that produced ultra-low gossypol cottonseed (ULGCS) [166]. Recently, ultra-low gossypol cottonseed (ULGCS) was obtained by using

PTGS and seed-specific promoter (α-globulin) through suppression of CDN genes and these lines are under field evaluations [167]. In the future, increased research investment on biotic and abiotic stresses through a transgenic approach is needed. Much focus is required for exploiting and improving cotton fiber and yield traits with the help of alien gene incorporation. In regards to public acceptance and questions, there is a need to carry on a massive public awareness campaign i.e. benefits, biosafety and risk assessment.

11. CRISPR Cas system for crop improvement in cotton

In the last decade, there is a revolution in the field of genome modification and continuous advancement in the targeted genome modification technologies. Genome editing tools like zinc-finger nucleases (ZFN), transcription activator-like effectors nucleases (TALENs) were extensively used before the advent of CRISPR Cas9 technology. These ZFN and TALENs technologies didn't become as popular as that of CRISPR Cas9 due to low efficiency, low specificity, low engineering feasibility and low design simplicity. CRISPR/Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated protein) system is the latest gene-editing technology, which has the power to alter the DNA or the code of life. Scientists predict that CRISPR has enormous potential for the next green revolution by 2050. To date, the CRISPR/Cas9 system has been successfully applied to efficient genome editing not only in model plants species but also in crop plant species. Many institutions and several groups all over the world are studying the feasibilities of the CRISPR/Cas9 system in cotton (*Gossypium hirsutum* L.). Successful use of CRISPR/Cas9 system in cotton still relies on Agrobacterium-mediated transformation and tissue culture, a genotype-dependent and low-efficiency process, but it provides a powerful tool for cotton functional genomics as it seems to be more efficient than RNA interference (RNAi) and virus-induced gene silencing (VIGS) in terms of knocking out the function of target genes [168]. The applications of the CRISPR/Cas9 system in cotton is been presented in **Table 6**. Chen and coworkers [169] demonstrated for the first time that the CRISPR/Cas9 can be used for advanced functional genomic research in cotton through targeted mutagenesis of two endogenous genes [Cloroplastos alterados 1 (GhCLA1) and vacuolar H + -pyrophosphatase (GhVP)]. Many factors influence the efficiency of the CRISPR/Cas9 system to obtain high mutation rates. [174] studied the sgRNA expression and mutagenesis efficiency by taking the endogenous U6 promoter over the existing one (Arabidopsis AtU6-29 promoter). Improved mutagenesis efficiency (4 to 6 times) was obtained by the use of an endogenous U6 promoter to drive the sgRNA expression. This study provided a fast and effective method to validate sgRNA mutagenesis efficiency in cotton using CRISPR/Cas9. Gao and coworkers [168] analyzed the nature of mutations induced by the CRISPR/Cas9 system through transient expression study of two genes Translation elongation factor 1 (GhEF1) and Phytoene desaturase (GhPDS) in cotton. The CRISPR/Cas9 system has been used for multiple sites targeting and simultaneously editing of multiple genes. Wang and his colleagues [171] successfully utilized the CRISPR/Cas9 system in allotetraploid cotton and accomplished multiple sites genome editing by targeting the exogenously transformed gene Discosoma red fluorescent protein2 (DsRed2) and an endogenous gene Cloroplastos alterados 1 (GhCLA1).

CRISPR/Cas9 has been used to edit a couple of agronomically important cotton genes, such as the genes involved in fiber development (GhMYB25-like A and GhMYB25-like D) [170] and a gene encoding arginase (ARG) for the increased lateral root formation [160]. Zhu and his colleagues [172] demonstrated the high

Sl. no.	Gene	Mutation type	Method of Cas9 system delivery	Phenotype	Gene function	Reference
1	GhCLA1 (Chloroplasts alterados 1)	Nucleotide insertion and substitution	Agrobacterium-mediated transformation	Albino phenotype was observed	A novel gene for chloroplast development	[169]
	GhVP (vacuolar H+-pyrophosphatase)			—	Involved in both acidify intracellular compartments and to transport protons across the plasma membrane.	
2	GhMYB25-like A & GhMYB25-like D	Nucleotide insertions and deletions (indels)	Agrobacterium-mediated transformation	—	GhMYB25-like is involved in the development of cotton fiber.	[170]
3	Arginase (ARG)	Nucleotide insertions and deletions (indels)	Agrobacterium-mediated transformation	Improved lateral root system	Plays an important role in the regulation of lateral root formation.	[171]
4	GhPDS, GhCLA1 & GhEF1	Deletions (64%)	Agrobacterium mediated transformation	Albino phenotypes observed	A novel gene for chloroplast development.	[168]
5	An endogenous gene GhCLA1 and DsRed2 (Discosoma red fluorescent protein2)	Nucleotide insertions and deletions (indels)	Agrobacterium tumefaciensmediated transformation	Disappeared red fluorescence and showed albino phenotype	AtCLA1 is involved in the development of chloroplast. DsRed2 protein is utilized as a reporter due to its different benefits over other report proteins.	[171]
6	ALARP	Nucleotide insertions and deletions	Agrobacterium-mediated transformation	—	A gene encoding alanine-rich protein that is preferentially expressed in cotton fibers	[172]
7	Cotton Gland Formation (*CGF3*)	Nucleotide insertions and deletions	Agrobacterium-mediated transformation	Glandless phenotype	plays a critical role in the formation of glands in the cotton plant	[162]
8	*Cotton Gland Pigmentation 1* (CGP1)	Nucleotide insertions and deletions	Agrobacterium-mediated transformation	Decreased accumulation of gossypol and related terpenoids, as well as the color intensity in glands	CGP1 is an MYB Transcription Factor that regulates gossypol accumulation but not gland morphogenesis.	[173]

Table 6.
*The applications of the CRISPR/Cas9 system in cotton (*Gossypium hirsutum *L.).*

editing efficiency of the CRISPR/Cas9 system in cotton by targeting-ALARP, a gene encoding an alanine-rich protein that is preferentially expressed in cotton fibers. CRISPR/Cas9 knockout of the Cotton Gland Formation (CGF3) gene resulted in a glandless phenotype in cotton. Gao and coworkers [168] confirmed the important role of Cotton Gland Pigmentation 1 (CGP1) in gland biology through CRISPR knockout of CGP1. Decreased accumulation of gossypol and of related terpenoids was observed in the CRISPR knockout plants. The above successful studies indicate that the CRISPR Cas9 system can further be effectively utilized in the functional genomics of cotton research. However, there are some limitations of the CRISPR/Cas9 system, including off-target effects, difficulties in PAM (protospacer adjacent motif) sequence selection for fewer potential target sites, and difficulties in generating homozygous mutations in the offspring [175–177]. Therefore, there is a lot of scope for the modification of the CRISPR Cas9 system or finding new alternative CRISPR systems. Zeng and co-workers [178], for the first time, established an efficient CRISPR/LbCpf1 system to expand the scope of genome editing in allotetraploid cotton by targeting the cotton endogenous gene Cloroplastos alterados (GhCLA). In addition to CRISPR/Cas9 & CRISPR/LbCpf1 system, a new effector with a single nuclease domain, a relatively small size, with low-frequency off-target effects, and cleavage capability under high temperature has been recently established and designated CRISPR/Cas12b (C2c1) [179]. CRISPR/Cas12b is a heat-induced system which requires a temperature ranging between 40 and 55°C for effective cleavage, when the temperature is lower than 40°C, cleavage cannot be accomplished [180, 181]. Recently the manipulation of the Cloroplastos alterados (GhCLA) gene in cotton plants using AacCas12b has been successfully established with no off-target effects. This system is ideal for plant species that can tolerate temperatures above 40°C, such as cotton that can grow well at temperatures reaching 45°C [171]. Some researchers are deactivating one or both of the Cas9's cutting domains and fusing new enzymes onto the protein. Cas9 can then be used to transport those enzymes to a specific DNA sequence. In one example, the Cas9 is fused to an enzyme, a deaminase, which mutates specific DNA bases eventually replacing cytidine with thymidine. [182] developed a new base editor system (GhBE3) consisting of a cytidine deaminase domain fused with nCas9 and uracil glycosylase inhibitor (UGI), for use in allotetraploid cotton, and obtained high base-editing efficiency with no detectable off-target effects. From all the above studies, it is indicated that CRISPR/Cas9 and its alternatives are potential gene-editing tools which would be superior to RNAi for cotton functional genomics. In future, this technology will have much scope for targeting tolerance to sucking pests, increased fiber yield and improved fiber quality traits in cotton.

12. Conclusion

Conventional Cotton breeders around the world had made a significant impact on cotton productivity improvement and germplasm conservation by their meticulous research that was developed during the early time is invaluable and highly significant. The advent of newer technology is an added advantage to the new young breeders since they can take advantage of newer technologies, but however prior knowledge of thorough conventional breeding, its limitations, and advantages of advanced molecular breeding, its limitation has to be kept in mind. A small mistake made initially during early molecular breeding may make us pay a heavy price in the end. The wealth of data like QTLs/Transgenic events/Gene-edited lines already developed can be cautiously used in crop improvement programs and further

research in advanced technologies like genomic selection, fine mapping, and gene editing has to be the priority area of research for the sustainable cotton production.

Conflict of interest

Nil.

Notes/thanks/other declarations

We thank sincerely: Dr. B. M. Khadi, former Director, ICAR- Central Institute for Cotton Research, Nagpur, (India).

Author details

Ishwarappa S. Katageri*, S. Anjan Gowda, Prashanth B.N, Mahesh Biradar, Rajeev M and Rajesh S. Patil
University of Agricultural Sciences, Dharwad, Karnataka, India

*Address all correspondence to: katageriis@uasd.in

References

[1] Shikha A, Bhuyan S. Cotton Crop: Various Aspects and Transition from Past, Present and Future. International Journal of Agriculture & Environmental Science. 2017;4.

[2] Moulherat C, Tengberg M, Haquet J-F, Mille Bt. First evidence of cotton at Neolithic Mehrgarh, Pakistan: analysis of mineralized fibres from a copper bead. Journal of Archaeological Science. 2002;29(12):1393-401.

[3] Krifa M, Stevens SS. Cotton utilization in conventional and non-conventional textiles—a statistical review. Agricultural Sciences. 2016;7(10):747-58.

[4] Mahir FI, Keya KN, Sarker B, Nahiun KM, Khan RA. A brief review on natural fiber used as a replacement of synthetic fiber in polymer composites. Materials Engineering Research. 2019;1(2):86-97.

[5] FAO I. Measuring Sustainability in Cotton Farming Systems: Towards a Guidance Framework. Rome (Accessed 05 July 18). 2015.

[6] Faostat. Food and Agriculture Organisation of the United Nations 2020. Available from: http://www.fao.org/faostat/en/#data.

[7] Ficci aWA. Global Shifts in Textile Industry & India's Position 2016. 1-29. Available from: http://ficci.in/spdocument/20817/3-FICCI-TAG-2016-Whitepaper.pdf.

[8] Lutz W, KC S. Dimensions of global population projections: what do we know about future population trends and structures? Philosophical Transactions of the Royal Society B: Biological Sciences. 2010;365(1554):2779-91.

[9] CICR. Vision 2050: Indian Council of Agricultural Research, New Delhi; 2015. Available from: https://www.cicr.org.in/.

[10] Chen ZJ, Scheffler BE, Dennis E, Triplett BA, Zhang T, Guo W, et al. Toward sequencing cotton (Gossypium) genomes. Plant physiology. 2007;145(4):1303-10.

[11] Wendel JF, Brubaker C, Alvarez I, Cronn R, Stewart JM. Evolution and natural history of the cotton genus. Genetics and genomics of cotton: Springer; 2009. p. 3-22.

[12] Wendel JF, Grover CE. Taxonomy and evolution of the cotton genus, Gossypium. Cotton. 2015;57:25-44.

[13] Craven L, Stewart JMD, Brown A, Grace J, editors. The Australian wild species of Gossypium. Challenging the Future: Proceedings of the World Cotton Research Conference I; 1994: CSIRO.

[14] Khadi B, Santhy V, Yadav M. Cotton: an introduction. Cotton: Springer; 2010. p. 1-14.

[15] Rapp RA, Alvarez I, Wendel JF. Molecular confirmation of the position of Gossypium trifurcatum Vollesen. Genetic Resources and Crop Evolution. 2005;52(6):749-53.

[16] Brubaker CL, Wendel JF. On the specific status of Gossypium lanceolatum Todaro. Genetic resources and crop evolution. 1993;40(3):165-70.

[17] Grover C, Zhu X, Grupp K, Jareczek J, Gallagher J, Szadkowski E, et al. Molecular confirmation of species status for the allopolyploid cotton species, Gossypium ekmanianum Wittmack. Genetic Resources and Crop Evolution. 2015;62(1):103-14.

[18] Gallagher JP, Grover CE, Rex K, Moran M, Wendel JF. A new species of

cotton from Wake Atoll, Gossypium stephensii (Malvaceae). Systematic Botany. 2017;42(1):115-23.

[19] Shim J, Mangat P, Angeles-Shim R. Natural variation in wild Gossypium species as a tool to broaden the genetic base of cultivated cotton. J Plant Sci Curr Res. 2018;2(005).

[20] Ware JO. Plant Breeding and the Cotton Industry: US Department of agriculture; 1936.

[21] Islam MS. Muslin! Our story: Drik Picture Library; 2016.

[22] Wendel JF, Brubaker CL, Percival AE. Genetic diversity in *Gossypium hirsutum* and the origin of upland cotton. American Journal of Botany. 1992;79(11):1291-310.

[23] Gumber R, Gill M, Dharminder Pathak GJ, Sarlach R. History and status of cotton. Cotton Research in Punjab. 2008:1-2.

[24] Manickam S, Gururajan K, Gopalakrishnan N. Development of isogenic restorer line in extra long staple cotton variety Suvin. Electronic Journal of Plant Breeding. 2010;1(4):632-6.

[25] CICR. Central Institute for Cotton research Nagpur2015. Available from: https://www.cicr.org.in/CropImprovement.html.

[26] Abdurakhmonov IY. World cotton germplasm resources: BoD–Books on Demand; 2014.

[27] Percy R, Frelichowski J, Arnold M, Campbell B, Dever J, Fang D, et al. The US National cotton germplasm collection—its contents, preservation, characterization, and evaluation. World cotton germplasm resources InTech, Rijeka. 2014:167-201.

[28] Jia Y, Sun J, Du X. Cotton germplasm resources in China. World cotton germplasm resources. 2014:35-53.

[29] Campbell B, Saha S, Percy R, Frelichowski J, Jenkins JN, Park W, et al. Status of the global cotton germplasm resources. Crop science. 2010;50(4):1161-79.

[30] Stiller WN, Wilson IW. Australian cotton germplasm resources. World Cotton Germplasm Resources. 2014:1-34.

[31] Jenkins JN, McCarty JC, Hayes RW, Stelly DM, Saha S. Registration of RMBHMTUP-C4, a Random-Mated Cotton Population Containing Alleles from Four Gossypium Species. Journal of Plant Registrations. 2019;13(3):411-5.

[32] Jenkins J, McCarty Jr J, Gutierrez O, Hayes R, Jones D. Registration of RMBUP-C4, a Random-Mated Population with *Gossypium barbadense* L. Alleles Introgressed into Upland Cotton Germplasm. Journal of Plant Registrations. 2013;7(2):224-8.

[33] Constable G, Reid P, Thomson N. Approaches utilized in breeding and development of cotton cultivars in Australia. 2001.

[34] Patel C. Evolution of hybrid-4 cotton. Current Science. 1981;50(8):343-6.

[35] Katarki B. Varalaxmi hybrid cotton--a valuable import substitute. Cotton Dev. 1972.

[36] Kadapa S, Malgar S, Onkarappa K. Variability for number of glands in the leaves of a few cotton varieties (*Gossypium hirsutum* L.). Current Research. 1980;9(7):121-2.

[37] Kadapa S, Thimmaiah G, editors. Breeding for Bollworm Resistance in Cotton (*G. hirsutum* L.). Proceedings of the IV International Congress of Genetics, New Delhi, Abs; 1983.

[38] Rao P, Kadappa S, Katageri I. Variability in morphological characters

of bollworm tolerant and susceptible cotton genotypes. Journal of Maharashtra Agricultural Universities (India). 1991.

[39] Katageri ISaK, S.N. Heterosis for yield and yield contributing characters in bollworm tolearant interspecific hybrids (*G. hirsutum* x *G. barbadense*). Indian J Genet. 1989;49(1):107-11.

[40] Katageri IS, Kadapa, S.N. and Goud, J.V. A case study of exploitation of heterosis using multiple pest tolerant *G. hirsutum* cotton strains. . Indian J Genet. 1990;50(4):314-9.

[41] Sundaram V. Handbook of cotton in India. 1999.

[42] Meredith W. Registration of 'MD51ne'cotton. Crop science. 1993;33(6):1415-.

[43] Meredith W. Registration of eight sub-okra, semi-smooth, and nectariless near-isolines of DES 119 cotton germplasm. Crop science. 1998;38(6):1725-.

[44] Ibragimov PS, Avtonomov VA, Amanturdiev AB, Namazov SE, Zaurov DE, Molnar TJ, et al. Uzbek scientific research institute of cotton breeding and seed production: breeding and germplasm resources. J Cotton Sci. 2008;12:62-72.

[45] Brandies P, Peel E, Hogg CJ, Belov K. The value of reference genomes in the conservation of threatened species. Genes. 2019;10(11):846.

[46] Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, Jin D, et al. Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres. Nature. 2012;492(7429):423-7.

[47] Huang G, Wu Z, Percy RG, Bai M, Li Y, Frelichowski JE, et al. Genome sequence of *Gossypium herbaceum* and genome updates of *Gossypium arboreum* and *Gossypium hirsutum* provide insights into cotton A-genome evolution. Nature genetics. 2020;52(5):516-24.

[48] Yang Z, Ge X, Yang Z, Qin W, Sun G, Wang Z, et al. Extensive intraspecific gene order and gene structural variations in upland cotton cultivars. Nature communications. 2019;10(1):1-13.

[49] Hu Y, Chen J, Fang L, Zhang Z, Ma W, Niu Y, et al. *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. Nature genetics. 2019;51(4):739-48.

[50] Chen ZJ, Sreedasyam A, Ando A, Song Q, De Santiago LM, Hulse-Kemp AM, et al. Genomic diversifications of five Gossypium allopolyploid species and their impact on cotton improvement. Nature genetics. 2020;52(5):525-33.

[51] Wang M, Tu L, Yuan D, Zhu D, Shen C, Li J, et al. Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*. Nature genetics. 2019;51(2):224-9.

[52] Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, et al. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. Nature biotechnology. 2015;33(5):531-7.

[53] Li F, Fan G, Lu C, Xiao G, Zou C, Kohel RJ, et al. Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. Nature biotechnology. 2015;33(5):524-30.

[54] Yuan D, Tang Z, Wang M, Gao W, Tu L, Jin X, et al. The genome sequence of Sea-Island cotton (*Gossypium barbadense*) provides insights into the allopolyploidization and development

of superior spinnable fibres. Scientific reports. 2015;5:17662.

[55] Du X, Huang G, He S, Yang Z, Sun G, Ma X, et al. Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits. Nature genetics. 2018;50(6):796-802.

[56] Li F, Fan G, Wang K, Sun F, Yuan Y, Song G, et al. Genome sequence of the cultivated cotton *Gossypium arboreum*. Nature genetics. 2014;46(6):567-72.

[57] Udall JA, Long E, Hanson C, Yuan D, Ramaraj T, Conover JL, et al. De novo genome sequence assemblies of Gossypium raimondii and Gossypium turneri. G3: Genes, Genomes, Genetics. 2019;9(10):3079-85.

[58] Wang K, Wang Z, Li F, Ye W, Wang J, Song G, et al. The draft genome of a diploid cotton Gossypium raimondii. Nature genetics. 2012;44(10):1098-103.

[59] Cai Y, Cai X, Wang Q, Wang P, Zhang Y, Cai C, et al. Genome sequencing of the Australian wild diploid species Gossypium australe highlights disease resistance and delayed gland morphogenesis. Plant biotechnology journal. 2020;18(3):814-28.

[60] Grover CE, Pan M, Yuan D, Arick MA, Hu G, Brase L, et al. The Gossypium longicalyx genome as a resource for cotton breeding and evolution. G3: Genes, Genomes, Genetics. 2020;10(5):1457-67.

[61] Shi Y-H, Zhu S-W, Mao X-Z, Feng J-X, Qin Y-M, Zhang L, et al. Transcriptome profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber cell elongation. The plant cell. 2006;18(3):651-64.

[62] Lu Q, Shi Y, Xiao X, Li P, Gong J, Gong W, et al. Transcriptome analysis suggests that chromosome introgression fragments from sea island cotton (*Gossypium barbadense*) increase fiber strength in upland cotton (*Gossypium hirsutum*). G3: Genes, Genomes, Genetics. 2017;7(10):3469-79.

[63] Bajwa KS, Shahid AA, Rao AQ, Bashir A, Aftab A, Husnain T. Stable transformation and expression of GhEXPA8 fiber expansin gene to improve fiber length and micronaire value in cotton. Frontiers in plant science. 2015;6:838.

[64] Padmalatha KV, Dhandapani G, Kanakachari M, Kumar S, Dass A, Patil DP, et al. Genome-wide transcriptomic analysis of cotton under drought stress reveal significant down-regulation of genes and pathways involved in fibre elongation and up-regulation of defense responsive genes. Plant molecular biology. 2012;78(3):223-46.

[65] Padmalatha KV, Patil DP, Kumar K, Dhandapani G, Kanakachari M, Phanindra ML, et al. Functional genomics of fuzzless-lintless mutant of *Gossypium hirsutum* L. cv. MCU5 reveal key genes and pathways involved in cotton fibre initiation and elongation. BMC genomics. 2012;13(1):1-15.

[66] Hande AS, Katageri IS, Jadhav MP, Adiger S, Gamanagatti S, Padmalatha KV, et al. Transcript profiling of genes expressed during fibre development in diploid cotton (*Gossypium arboreum* L.). BMC genomics. 2017;18(1):675.

[67] Li P-t, Wang M, Lu Q-w, Ge Q, Liu A-y, Gong J-w, et al. Comparative transcriptome analysis of cotton fiber development of Upland cotton (*Gossypium hirsutum*) and Chromosome Segment Substitution Lines from *G. hirsutum* × *G. barbadense*. BMC genomics. 2017;18(1):705.

[68] Yu S, Cheng G, Zhang L, Wei H, Wang H, Lu J. Transcriptome analysis reveals gene expression associated with fuzz fiber initiation regulated by high-temperature in *Gossypium barbadense*. 2020.

[69] Malik W, Khan AA, Cheema HMN, Aslam U, Iqbal MZ, Qayyum A, et al. Transcriptome analysis of pigment related genes in colored cotton. International Journal of Agriculture and Biology. 2015;17(1):205-10.

[70] Tang Z, Fan Y, Zhang L, Zheng C, Chen A, Sun Y, et al. Quantitative metabolome and transcriptome analysis reveals complex regulatory pathway underlying photoinduced fiber color formation in cotton. Gene. 2020:145180.

[71] Gong W, He S, Tian J, Sun J, Pan Z, Jia Y, et al. Comparison of the transcriptome between two cotton lines of different fiber color and quality. PloS one. 2014;9(11):e112966.

[72] Han M, Lu X, Yu J, Chen X, Wang X, Malik WA, et al. Transcriptome Analysis Reveals Cotton (*Gossypium hirsutum*) Genes That Are Differentially Expressed in Cadmium Stress Tolerance. International journal of molecular sciences. 2019;20(6):1479.

[73] Li Z-B, Zeng X-Y, Xu J-W, Zhao R-H, Wei Y-N. Transcriptomic profiling of cotton *Gossypium hirsutum* challenged with low-temperature gradients stress. Scientific data. 2019;6(1):1-7.

[74] Bowman MJ, Park W, Bauer PJ, Udall JA, Page JT, Raney J, et al. RNA-Seq transcriptome profiling of upland cotton (*Gossypium hirsutum* L.) root tissue under water-deficit stress. PLoS One. 2013;8(12):e82634.

[75] Li R, Rashotte AM, Singh NK, Lawrence KS, Weaver DB, Locy RD. Transcriptome analysis of cotton (*Gossypium hirsutum* L.) genotypes that are susceptible, resistant, and hypersensitive to reniform nematode (Rotylenchulus reniformis). PLoS One. 2015;10(11):e0143261.

[76] Curtiss J, Rodriguez-Uribe L, Stewart JM, Zhang J. Identification of differentially expressed genes associated with semigamy in Pima cotton (*Gossypium barbadense*l.) through comparative microarray analysis. BMC plant biology. 2011;11(1):49.

[77] Naqvi RZ, Zaidi SS-e-A, Mukhtar MS, Amin I, Mishra B, Strickler S, et al. Transcriptomic analysis of cultivated cotton *Gossypium hirsutum* provides insights into host responses upon whitefly-mediated transmission of cotton leaf curl disease. PloS one. 2019;14(2):e0210011.

[78] Wang L, Yin Y, Wang L-F, Wang M, Zhao M, Tian Y, et al. Transcriptome profiling of the elongating internode of cotton (*Gossypium hirsutum* L.) seedlings in response to mepiquat chloride. Frontiers in plant science. 2019;10:1751.

[79] Shreeraksha R. J. ISK, S. A. Gowda and N. V. M. Kumar. DNA methylation in differential gene expression during fiber initiation of cotton (Gossypium spp.). Journal of Farm science. 2020;33(3):319-25.

[80] Zhang M, Guo L, Qi T, Zhang X, Tang H, Wang H, et al. Integrated Methylome and Transcriptome Analysis between the CMS-D2 Line ZBA and Its Maintainer Line ZB in Upland Cotton. International journal of molecular sciences. 2019;20(23):6070.

[81] Fan HH, Wei J, Li TC, Li ZP, Guo N, Cai YP, et al. DNA methylation alterations of upland cotton (*Gossypium hirsutum*) in response to cold stress. Acta physiologiae plantarum. 2013;35(8):2445-53.

[82] Wang B, Zhang M, Fu R, Qian X, Rong P, Zhang Y, et al.

Epigenetic mechanisms of salt tolerance and heterosis in Upland cotton (*Gossypium hirsutum* L.) revealed by methylation-sensitive amplified polymorphism analysis. Euphytica. 2016;208(3):477-91.

[83] Sun Q, Qiao J, Zhang S, He S, Shi Y, Yuan Y, et al. Changes in DNA methylation assessed by genomic bisulfite sequencing suggest a role for DNA methylation in cotton fruiting branch development. PeerJ. 2018;6:e4945.

[84] Reinisch AJ, Dong J-M, Brubaker CL, Stelly DM, Wendel JF, Paterson AH. A detailed RFLP map of cotton, *Gossypium hirsutum* x *Gossypium barbadense*: chromosome organization and evolution in a disomic polyploid genome. Genetics. 1994;138(3):829-47.

[85] Kohel RJ, Yu J, Park Y-H, Lazo GR. Molecular mapping and characterization of traits controlling fiber quality in cotton. Euphytica. 2001;121(2):163-72.

[86] Lacape J-M, Nguyen T-B, Thibivilliers S, Bojinov B, Courtois B, Cantrell RG, et al. A combined RFLP SSR AFLP map of tetraploid cotton based on a *Gossypium hirsutum*× *Gossypium barbadense* backcross population. Genome. 2003;46(4):612-26.

[87] Lacape J-M, Jacobs J, Arioli T, Derijcker R, Forestier-Chiron N, Llewellyn D, et al. A new interspecific, *Gossypium hirsutum*× *G. barbadense*, RIL population: towards a unified consensus linkage map of tetraploid cotton. Theoretical and applied genetics. 2009;119(2):281-92.

[88] Yu Y, Yuan D, Liang S, Li X, Wang X, Lin Z, et al. Genome structure of cotton revealed by a genome-wide SSR genetic map constructed from a BC 1 population between *Gossypium hirsutum* and *G. barbadense*. BMC genomics. 2011;12(1):1-14.

[89] Lin Z-x, He D, Zhang X-l, Nie Y, Guo X, Feng C, et al. Linkage map construction and mapping QTL for cotton fibre quality using SRAP, SSR and RAPD. Plant breeding. 2005;124(2):180-7.

[90] Zhang Z-S, Xiao Y-H, Luo M, Li X-B, Luo X-Y, Hou L, et al. Construction of a genetic linkage map and QTL analysis of fiber-related traits in upland cotton (*Gossypium hirsutum* L.). Euphytica. 2005;144(1-2):91-9.

[91] Hu J, Vick BA. Target region amplification polymorphism: a novel marker technique for plant genotyping. Plant Molecular Biology Reporter. 2003;21(3):289-94.

[92] Liu B, Wendel JF. Intersimple sequence repeat (ISSR) polymorphisms as a genetic marker system in cotton. Molecular Ecology Notes. 2001 Sep;1(3):205-8

[93] Sekhar L, Khadi B, Patil RS, Katageri I, Mukri G. Biochemical and molecular dissection of thermo-sensitive genetic male sterility in diploid cotton (*Gossypium arboreum* L.). Journal of Environmental Biology. 2016;37(4):579.

[94] Kumar NM, Katageri IS, Gowda SA, Adiger S, Yadava SK, Lachagari VR. 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.

[95] Handi SS, Katageri IS, Adiger S, Jadhav MP, Lekkala SP, Reddy Lachagari VB. Association mapping for seed cotton yield, yield components and fibre quality traits in upland cotton (*Gossypium hirsutum* L.) genotypes. Plant Breeding. 2017;136(6):958-68.

[96] Ramesh U, Methre R, Kumar N, Katageri IS, Gowda SA, Adiger S, et al. Genome mapping and molecular markers identification for yield, yield component and fibre quality traits in tetraploid cotton. Plant Breeding. 2019;138(6):880-96.

[97] Sankeshwar M, Jadhav M, Adiger S, Patil RS, Katageri I. Mapping of QTLs for traits related to leaf pubescence, jassid resistance and yield in cotton (Gossypium spp.). Indian J Genet. 2018;78(2):252-60.

[98] Yu J, Jung S, Cheng C-H, Ficklin SP, Lee T, Zheng P, et al. CottonGen: a genomics, genetics and breeding database for cotton research. Nucleic acids research. 2014;42(D1):D1229-D36.

[99] Hulse-Kemp AM, Lemm J, Plieske J, Ashrafi H, Buyyarapu R, Fang DD, et al. Development of a 63K SNP array for cotton and high-density mapping of intraspecific and interspecific populations of Gossypium spp. G3: Genes, Genomes, Genetics. 2015;5(6):1187-209.

[100] Cai C, Zhu G, Zhang T, Guo W. High-density 80 K SNP array is a powerful tool for genotyping *G. hirsutum* accessions and genome analysis. BMC genomics. 2017;18(1):654.

[101] Lee G-A, Koh H-J, Chung H-K, Dixit A, Chung J-W, Ma K-H, et al. Development of SNP-based CAPS and dCAPS markers in eight different genes involved in starch biosynthesis in rice. Molecular Breeding. 2009;24(1):93-101.

[102] Lestari P, Koh HJ. Development of new CAPS/dCAPS and SNAP markers for rice eating quality. HAYATI Journal of Biosciences. 2013;20(1):15-23.

[103] Obala J, Saxena RK, Singh VK, Kumar CS, Saxena K, Tongoona P, et al. Development of sequence-based markers for seed protein content in pigeonpea. Molecular Genetics and Genomics. 2019;294(1):57-68.

[104] Li P, Kirungu JN, Lu H, Magwanga RO, Lu P, Cai X, et al. SSR-Linkage map of interspecific populations derived from Gossypium trilobum and *Gossypium thurberi* and determination of genes harbored within the segregating distortion regions. PloS one. 2018;13(11):e0207271.

[105] Zhang K, Kuraparthy V, Fang H, Zhu L, Sood S, Jones DC. 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. 2019;20(1):889.

[106] Ma J, Pei W, Ma Q, Geng Y, Liu G, Liu J, et al. QTL analysis and candidate gene identification for plant height in cotton based on an interspecific backcross inbred line population of *Gossypium hirsutum* × *Gossypium barbadense*. Theoretical and Applied Genetics. 2019;132(9):2663-76.

[107] Naoumkina M, Thyssen GN, Fang DD, Jenkins JN, McCarty JC, Florane CB. Genetic and transcriptomic dissection of the fiber length trait from a cotton (*Gossypium hirsutum* L.) MAGIC population. BMC genomics. 2019;20(1):112.

[108] Wang W, Sun Y, Yang P, Cai X, Yang L, Ma J, et al. A high density SLAF-seq SNP genetic map and QTL for seed size, oil and protein content in upland cotton. BMC genomics. 2019;20(1):599.

[109] Zhang C, Li L, Liu Q, Gu L, Huang J, Wei H, et al. Identification of loci and candidate genes responsible for fiber length in upland cotton (*Gossypium hirsutum* L.) via association

mapping and linkage analyses. Frontiers in plant science. 2019;10:53.

[110] Zhang T, Yuan Y, Yu J, Guo W, Kohel RJ. Molecular tagging of a major QTL for fiber strength in Upland cotton and its marker-assisted selection. Theoretical and Applied Genetics. 2003;106(2):262-8.

[111] Rong J, Abbey C, Bowers JE, Brubaker CL, Chang C, Chee PW, et al. A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (Gossypium). Genetics. 2004;166(1):389-417.

[112] Brubaker CL, Brown AH. The use of multiple alien chromosome addition aneuploids facilitates genetic linkage mapping of the Gossypium G genome. Genome. 2003;46(5):774-91.

[113] Chen H, Khan MKR, Zhou Z, Wang X, Cai X, Ilyas MK, et al. A high-density SSR genetic map constructed from a F2 population of *Gossypium hirsutum* and Gossypium darwinii. Gene. 2015;574(2):273-86.

[114] Wang B, Zhuang Z, Zhang Z, Draye X, Shuang L-S, Shehzad T, et al. Advanced Backcross QTL Analysis of Fiber Strength and Fineness in a Cross between *Gossypium hirsutum* and G. mustelinum. Frontiers in plant science. 2017;8:1848.

[115] Kushanov FN, Buriev ZT, Shermatov SE, Turaev OS, Norov TM, Pepper AE, et al. QTL mapping for flowering-time and photoperiod insensitivity of cotton Gossypium darwinii Watt. PloS one. 2017;12(10):e0186240.

[116] Kirungu JN, Deng Y, Cai X, Magwanga RO, Zhou Z, Wang X, et al. Simple sequence repeat (SSR) genetic linkage map of D genome diploid cotton derived from an interspecific cross between Gossypium davidsonii and Gossypium klotzschianum. International journal of molecular sciences. 2018;19(1):204.

[117] Keerio AA, Shen C, Nie Y, Ahmed MM, Zhang X, Lin Z. QTL mapping for fiber quality and yield traits based on introgression lines derived from *Gossypium hirsutum* × *G. tomentosum*. International journal of molecular sciences. 2018;19(1):243.

[118] Magwanga RO, Lu P, Kirungu JN, Diouf L, Dong Q, Hu Y, 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. International journal of molecular sciences. 2018;19(6):1614.

[119] Zheng J, Oluoch G, Riaz MK, Wang X, Cai X, Zhou Z, et al. Mapping QTLs for drought tolerance in an F2: 3 population from an inter-specific cross between *Gossypium tomentosum* and *Gossypium hirsutum*. Genetics and molecular research: GMR. 2016;15(3).

[120] Mekonnen T, Haileselassie T, Tesfaye K. Identification, mapping and pyramiding of genes/quantitative trait loci (qtls) for durable resistance of crops to biotic stresses. Journal of Plant Pathology and Microbiology. 2017;8(6).

[121] Mao G, Wei H, Hu W, Ma Q, Zhang M, Wang H, et al. Fine mapping and molecular characterization of the virescent gene vsp in Upland cotton (*Gossypium hirsutum*). Theoretical and Applied Genetics. 2019;132(7):2069-86.

[122] Mao G, Ma Q, Wei H, Su J, Wang H, Ma Q, et al. Fine mapping and candidate gene analysis of the virescent gene v 1 in Upland cotton (*Gossypium hirsutum*). Molecular genetics and genomics. 2018;293(1):249-64.

[123] Feng X, Cheng H, Zuo D, Zhang Y, Wang Q, Liu K, et al. Fine mapping

and identification of the fuzzless gene GaFzl in DPL972 (*Gossypium arboreum*). Theoretical and Applied Genetics. 2019;132(8):2169-79.

[124] Xu P, Gao J, Cao Z, Chee PW, Guo Q, Xu Z, et al. Fine mapping and candidate gene analysis of qFL-chr1, a fiber length QTL in cotton. Theoretical and Applied Genetics. 2017;130(6):1309-19.

[125] Fang X, Liu X, Wang X, Wang W, Liu D, Zhang J, et al. Fine-mapping qFS07. 1 controlling fiber strength in upland cotton (*Gossypium hirsutum* L.). Theoretical and applied genetics. 2017;130(4):795-806.

[126] Andres RJ, Bowman DT, Kaur B, Kuraparthy V. Mapping and genomic targeting of the major leaf shape gene (L) in Upland cotton (*Gossypium hirsutum* L.). Theoretical and applied genetics. 2014;127(1):167-77.

[127] Kumar P, He Y, Singh R, Davis RF, Guo H, Paterson AH, et al. Fine mapping and identification of candidate genes for a QTL affecting Meloidogyne incognita reproduction in Upland cotton. BMC genomics. 2016;17(1):567.

[128] Bernardo R, Yu J. Prospects for genomewide selection for quantitative traits in maize. Crop Science. 2007;47(3):1082-90.

[129] Nakaya A, Isobe SN. Will genomic selection be a practical method for plant breeding? Annals of botany. 2012;110(6):1303-16.

[130] Hayes B, Goddard M. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 2001;157(4):1819-29.

[131] Heffner EL, Sorrells ME, Jannink JL. Genomic selection for crop improvement. Crop Science. 2009;49(1):1-12.

[132] Shikha M, Kanika A, Rao AR, Mallikarjuna MG, Gupta HS, Nepolean T. Genomic selection for drought tolerance using genome-wide SNPs in maize. Frontiers in plant science. 2017;8:550.

[133] Vélez-Torres M, García-Zavala JJ, Hernández-RodríguezM,Lobato-OrtizR, López-Reynoso JJ, Benítez-Riquelme I, et al. Genomic prediction of the general combining ability of maize lines (*Zea mays* L.) and the performance of their single crosses. Plant Breeding. 2018;137(3):379-87.

[134] Liu X, Hu X, Li K, Liu Z, Wu Y, Wang H, et al. Genetic mapping and genomic selection for maize stalk strength. BMC plant biology. 2020;20:1-16.

[135] Guo R, Dhliwayo T, Mageto EK, Palacios-Rojas N, Lee M, Yu D, et al. Genomic Prediction of Kernel Zinc Concentration in Multiple Maize Populations Using Genotyping-by-Sequencing and Repeat Amplification Sequencing Markers. Frontiers in Plant Science. 2020;11.

[136] Spindel J, Begum H, Akdemir D, Virk P, Collard B, Redona E, et al. Genomic selection and association mapping in rice (*Oryza sativa*): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. PLoS Genet. 2015;11(2):e1004982.

[137] Xu Y, Wang X, Ding X, Zheng X, Yang Z, Xu C, et al. Genomic selection of agronomic traits in hybrid rice using an NCII population. Rice. 2018;11(1):1-10.

[138] Cui Y, Li R, Li G, Zhang F, Zhu T, Zhang Q, et al. Hybrid breeding of rice via genomic selection. Plant biotechnology journal. 2020;18(1):57-67.

[139] Yu S, Ali J, Zhang C, Li Z, Zhang Q. Genomic breeding of green super rice varieties and their deployment in Asia and Africa. Theoretical and Applied Genetics. 2020:1-16.

[140] Crossa J, de Los Campos G, Pérez P, Gianola D, Burgueño J, Araus JL, et al. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics. 2010;186(2):713-24.

[141] Dong H, Wang R, Yuan Y, Anderson J, Pumphrey M, Zhang Z, et al. Evaluation of the potential for genomic selection to improve spring wheat resistance to Fusarium head blight in the Pacific Northwest. Frontiers in plant science. 2018;9:911.

[142] Norman A, Taylor J, Edwards J, Kuchel H. Optimising genomic selection in wheat: Effect of marker density, population size and population structure on prediction accuracy. G3: Genes, Genomes, Genetics. 2018;8(9):2889-99.

[143] Veenstra LD, Poland J, Jannink JL, Sorrells ME. Recurrent genomic selection for wheat grain fructans. Crop Science. 2020.

[144] Lozada DN, Ward BP, Carter AH. Gains through selection for grain yield in a winter wheat breeding program. PLoS One. 2020;15(4):e0221603.

[145] de Oliveira AA, Pastina MM, da Costa Parrella RA, Noda RW, Simeone MLF, Schaffert RE, et al. Genomic prediction applied to high-biomass sorghum for bioenergy production. Molecular Breeding. 2018;38(4):49.

[146] Hunt CH, van Eeuwijk FA, Mace ES, Hayes BJ, Jordan DR. Development of genomic prediction in sorghum. Crop Science. 2018;58(2):690-700.

[147] Sapkota S, Boatwright JL, Jordan KE, Boyles R, Kresovich S. Multi-trait regressor stacking increased genomic prediction accuracy of sorghum grain composition. bioRxiv. 2020.

[148] Islam MS, Fang DD, Jenkins JN, Guo J, McCarty JC, Jones DC. Evaluation of genomic selection methods for predicting fiber quality traits in Upland cotton. Molecular Genetics and Genomics. 2020;295(1):67-79.

[149] Lenz PR, Beaulieu J, Mansfield SD, Clément S, Desponts M, Bousquet J. Factors affecting the accuracy of genomic selection for growth and wood quality traits in an advanced-breeding population of black spruce (*Picea mariana*). BMC genomics. 2017;18(1):335.

[150] Gapare W, Liu S, Conaty W, Zhu Q-H, Gillespie V, Llewellyn D, et al. Historical datasets support genomic selection models for the prediction of cotton fiber quality phenotypes across multiple environments. G3: Genes, Genomes, Genetics. 2018;8(5):1721-32.

[151] Sunilkumar G, Rathore KS. Transgenic cotton: factors influencing Agrobacterium-mediated transformation and regeneration. Molecular Breeding. 2001;8(1):37-52.

[152] Prashanth S, Katageri I, Vamadevaih H, Khadi B. Effect of external damage on regeneration of cotton explants (*Gossypium arboreum* and *G. barbadense*). Karnataka Journal of Agricultural Sciences. 2011;24(5):629-32.

[153] Jadhav M, Katageri I. Agrobacterium tumefaciens Mediated Genetic Transformation in Coker-312 (*Gossypium hirsutum* L.) Using Hypocotyls Explants. Int J Curr Microbiol App Sci. 2017;6(12):2771-9.

[154] Klein TM, Wolf ED, Wu R, Sanford J. High-velocity microprojectiles

for delivering nucleic acids into living cells. Nature. 1987;327(6117):70-3.

[155] Datta SK, Peterhans A, Datta K, Potrykus I. Genetically engineered fertile indica-rice recovered from protoplasts. Bio/technology. 1990;8(8):736-40.

[156] Fromm M, Taylor LP, Walbot V. Expression of genes transferred into monocot and dicot plant cells by electroporation. Proceedings of the National Academy of Sciences. 1985;82(17):5824-8.

[157] Leelavathi S, Sunnichan V, Kumria R, Vijaykanth G, Bhatnagar R, Reddy V. A simple and rapid Agrobacterium-mediatedtransformation protocol for cotton (*Gossypium hirsutum* L.): embryogenic calli as a source to generate large numbers of transgenic plants. Plant cell reports. 2004;22(7):465-70.

[158] Sumithra S, Katageri I, Vamadevaiah H. Factors influencing regeneration and Agrobacterium mediated transformation of Gossypium herbaceaum and *Gossypium hirsutum* cotton genotypes. Karnataka Journal of Agricultural Sciences. 2010;23(2):222-6.

[159] John ME, Stewart JM. Genes for jeans: biotechnological advances in cotton. Trends in biotechnology. 1992;10:165-70.

[160] Mao Y-B, Cai W-J, Wang J-W, Hong G-J, Tao X-Y, Wang L-J, et al. Silencing a cotton bollworm P450 monooxygenase gene by plant-mediated RNAi impairs larval tolerance of gossypol. Nature biotechnology. 2007;25(11):1307-13.

[161] Li B, Yang Y, Hu WR, Li XD, Cao JQ, Fan L. Over-expression of Gh UGP 1 in upland cotton improves fibre quality and reduces fibre sugar content. Plant Breeding. 2015;134(2):197-202.

[162] Jiang Y, Guo W, Zhu H, Ruan YL, Zhang T. Overexpression of GhSusA1 increases plant biomass and improves cotton fiber yield and quality. Plant biotechnology journal. 2012;10(3):301-12.

[163] Li F, Wu S, Lü F, Chen T, Ju M, Wang H, et al. Modified fiber qualities of the transgenic cotton expressing a silkworm fibroin gene. Chinese science bulletin. 2009;54(7):1210-6.

[164] Abdurakhmonov IY, Buriev ZT, Saha S, Jenkins JN, Abdukarimov A, Pepper AE. Phytochrome RNAi enhances major fibre quality and agronomic traits of the cotton *Gossypium hirsutum* L. Nature Communications. 2014;5(1):1-10.

[165] Benedict CR, Martin GS, Liu J, Puckhaber L, Magill CW. Terpenoid aldehyde formation and lysigenous gland storage sites in cotton: variant with mature glands but suppressed levels of terpenoid aldehydes. *Phytochemistry*. 2004 May 1;65(10):1351-9.

[166] Rathore KS, Sundaram S, Sunilkumar G, Campbell LM, Puckhaber L, Marcel S, et al. Ultra-low gossypol cottonseed: generational stability of the seed-specific, RNAi-mediated phenotype and resumption of terpenoid profile following seed germination. Plant biotechnology journal. 2012;10(2):174-83.

[167] Palle SR, Campbell LM, Pandeya D, Puckhaber L, Tollack LK, Marcel S, et al. RNA i-mediated Ultra-low gossypol cottonseed trait: performance of transgenic lines under field conditions. Plant biotechnology journal. 2013;11(3):296-304.

[168] Gao W, Long L, Tian X, Xu F, Liu J, Singh PK, et al. Genome editing in cotton with the CRISPR/Cas9 system. Frontiers in plant science. 2017;8:1364.

[169] Chen X, Lu X, Shu N, Wang S, Wang J, Wang D, et al. Targeted

mutagenesis in cotton (*Gossypium hirsutum* L.) using the CRISPR/Cas9 system. Scientific reports. 2017;7:44304.

[170] Li C, Unver T, Zhang B. A high-efficiency CRISPR/Cas9 system for targeted mutagenesis in cotton (*Gossypium hirsutum* L.). Scientific Reports. 2017;7:43902.

[171] Wang P, Zhang J, Sun L, Ma Y, Xu J, Liang S, et al. High efficient multisites genome editing in allotetraploid cotton (*Gossypium hirsutum*) using CRISPR/Cas9 system. Plant biotechnology journal. 2018;16(1):137-50.

[172] Zhu S, Yu X, Li Y, Sun Y, Zhu Q, Sun J. Highly efficient targeted gene editing in upland cotton using the CRISPR/Cas9 system. International journal of molecular sciences. 2018;19(10):3000.

[173] Gao W, Xu FC, Long L, Li Y, Zhang JL, Chong L, et al. The gland localized CGP1 controls gland pigmentation and gossypol accumulation in cotton. Plant biotechnology journal. 2020;18(7):1573.

[174] Long L, Guo D-D, Gao W, Yang W-W, Hou L-P, Ma X-N, et al. Optimization of CRISPR/Cas9 genome editing in cotton by improved sgRNA expression. Plant Methods. 2018;14(1):85.

[175] Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA, Liu DR. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nature biotechnology. 2013;31(9):839-43.

[176] Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nature protocols. 2013;8(11):2281-308.

[177] Zeng D, Li X, Huang J, Li Y, Cai S, Yu W, et al. Engineered Cas9 variant tools expand targeting scope of genome and base editing in rice. Plant Biotechnology Journal. 2020;18(6):1348-50.

[178] Li B, Rui H, Li Y, Wang Q, Alariqi M, Qin L, et al. Robust CRISPR/Cpf1 (Cas12a)-mediated genome editing in allotetraploid cotton (*Gossypium hirsutum*). Plant biotechnology journal. 2019;17(10):1862.

[179] Liu L, Chen P, Wang M, Li X, Wang J, Yin M, et al. C2c1-sgRNA complex structure reveals RNA-guided DNA cleavage mechanism. Molecular cell. 2017;65(2):310-22.

[180] Shmakov S, Abudayyeh OO, Makarova KS, Wolf YI, Gootenberg JS, Semenova E, et al. Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. Molecular cell. 2015;60(3):385-97.

[181] Yang H, Gao P, Rajashankar KR, Patel DJ. PAM-dependent target DNA recognition and cleavage by C2c1 CRISPR-Cas endonuclease. Cell. 2016;167(7):1814-28. e12.

[182] Qin L, Li J, Wang Q, Xu Z, Sun L, Alariqi M, et al. High-efficient and precise base editing of C• G to T• A in the allotetraploid cotton (*Gossypium hirsutum*) genome using a modified CRISPR/Cas9 system. Plant biotechnology journal. 2020;18(1):45-56.