December 2019
Volume XXXVII, No. 4
ISSN 1022-6303

The ICAC Recorder

International Cotton Advisory Committee

Table of Contents

The ICAC Recorder

Editorial

This volume of the ICAC RECORDER is a special issue focussing on 'Cotton Breeding and Genomics'. Plant breeding plays a key role in crop improvement. In modern times, genetics, genomics and molecular biology dominate agricultural research, often overriding the science of plant breeding. In a private conversation I had during a poster session at the recent Beltwide conference in Austin, Texas, a PhD student remarked to me that plant breeding was no longer 'sexy' and it was far less attractive than genomics and molecular sciences. It was interesting to hear him tell me confidently that 'plant breeding was a supporting science as it supports the science of genomics' and that genomics was the main gateway to improving agriculture — a generational change indeed. I always thought that it was the other way around where genetics, genomics and molecular sciences assisted and supported plant breeding, which played a pivotal role in crop improvement.

Each of these branches of cotton science has its own importance. Though plant breeding is widely believed to have contributed more substantially to crop improvement than any other branch of agricultural sciences, young researchers — for reasons best known to them — appear to be oriented more towards genomics and molecular biology to enhance fibre quality and increase yields. Nevertheless, plant breeding continues to be acknowledged for its critical role in crop improvement. At the ICAC Plenary Meeting in India in 2015, Dr Greg Constable, the renowned Australian cotton agronomist and plant breeder, pointed out that in Australia, plant breeding was responsible for 48% of the large yield gains of over 1,300 kg/ha of lint over a 30-year period; about 24% of the gain was due to modern crop management; and about 28% of the yield improvement was new cultivars responding more to modern management, thus attributing 76% of yield gains to the role played by cultivars. Interestingly, not once did he mention that transgenic technology may have played a role in yield enhancement in Australia. Dr Constable also showed that the theoretical yield was about 5,034 kg/ha of lint, while the best irrigated crops in Australia reached 3,500 kg/ha lint in 2015. Commendably, the Australian plant breeders not only succeeded in developing cultivars for high yields but also achieved high-quality fibre simultaneously, something that many plant breeders in other parts of the world think is not easy to achieve. In one of this issue's articles, Dr Shreekant Patil expresses his confidence that it is possible to develop cotton cultivars for high yields with high fibre quality through innovative approaches of plant breeding based on the principles of quantitative genetics.

Agricultural sciences are very complex and so are the concepts related to yield enhancement. High yields and high fibre quality are a function of many factors that include climate, water, nutrients, management and cultivars. A cultivar can contribute to high yields if it has the genetic capability to perform well under both ideal and challenging environments. Such cultivars are generally developed by plant breeders who have a sound background in agronomy, physiology and plant protection. What kind of plant-type should breeders develop for high yields? Big plants or small plants? Should they have100 bolls per plant or 10 bolls per plant? Short duration or long duration? Big seeded or small seeded? Single picking or multiple picking?

These questions solicit different responses from different countries. Plant breeders from India and Africa believe that to get high yields, it is important to have more bolls per plant, big-seeded cultivars, large-statured plants, long duration and multiple pickings. Plant breeders from Australia, China, Brazil, Turkey and USA, on the other hand, have strived over the past 30 to 40 years to achieve exactly the opposite of what the Indian and African plant breeders aim to develop. They obtained high yields and high-quality fibre with short-statured plants, fewer (10-15) bolls per plant, small-seeded, short-season cultivars and fewer pickings — and they also achieved a high harvest index in the process. Having diametrically opposed ideas seeking to achieve the same goal, high yields, is indeed perplexing. Nevertheless, the fact remains that with concepts of big plants, big seeds, more bolls per plant, long-season and multiple pickings, India and Africa aren't succeeding and are stuck with low annual national average yields of less than 500 kg lint per hectare whereas Australia, China, Turkey, Brazil harvest more than 1500 kg lint per hectare.

Australia's successful cotton journey is inspiring the world. It remains to be seen if cotton scientists of Africa and India are willing to be inspired by the Australian cotton success story. On the sidelines of the recent ICAC Plenary Meeting in Brisbane, I was part of an international group that visited cotton fields in Australia. The effects of drought were telling on the fields and the plants looked smaller than usual. '*Beyond doubt, this crop will give low yields*' declared a

The ICAC RECORDER (ISSN 1022-6303) is published four times a year by the Secretariat of the International Cotton Advisory Committee, 1629 K Street, NW, Suite 702, Washington, DC 20006-1636, USA. Editor: Keshav Kranthi <keshav@icac.org>. Subscription rate: $220.00 hard copy.

visitor from India. A cotton scientist from Africa seemed very unhappy with the 50-days old short-statured plants. He explained animatedly to the Australian farmer that the crop in his country would have reached waist height at 50 days and wouldn't be at half-knee height, as it was in Australia. He asserted that something was not right with the crop in Australia. It was interesting to hear the cool, casual response of the Australian farmer: '*Well, I usually get 11 to 12 bales from this kind of a crop and am confident of the same this year, too*'. The farmer was expecting 11 to 12 Australian bales (227 Kg per bale), which would translate to 2497 to 2724 kg lint per hectare — something that neither the African scientist nor the Indian delegate could dream of achieving in their countries. I feel that cotton scientists can greatly benefit by understanding the Australian approaches and sustainable strategies of yield enhancement through plant breeding and scientific crop management. There is also a need to introspect if the current plant breeding strategies in India and Africa would succeed in breaking yield stagnation or if a paradigm shift in plant breeding objectives would be necessary.

This issue of the ICAC RECORDER has articles that discuss plant breeding strategies to enhance yields. It has articles authored by Dr. Shreekant Patil, an eminent cotton geneticist and a renowned Indian cotton breeder; Dr. Dharminder Pathak, an acknowledged cotton breeder; and Prof. Yuolu Yuan, a cotton genomics expert from China. The three articles describe concepts, ideas and innovations in cotton breeding, genetics, genomics and molecular sciences in the context of the recent progress made in cotton cultivar improvement, which I earnestly believe will be very useful for cotton researchers to formulate projects for high yields and premium fibre quality. I am thankful to Dr H. B. Santosh, the young talented plant breeder of ICAR-Central Institute for Cotton Research, Nagpur India for sharing his critical insights on the articles.

-Keshav R Kranthi

Dr. JOHN YU -COTTON RESEARCHER OF THE YEAR 2019

Dr John Yu has been working with the USDA-ARS since 1995. He earned his PhD in plant genomics and molecular genetics at Cornell University. He demonstrated outstanding dedication, with 36 years of professional experience, and received significant recognition as an international authority in cotton genomics for developing genetic resources that are currently being used by scientists worldwide. He made major contributions on cotton genomics and germplasm, including genome mapping and sequencing, characterisation of gene pools, and identification of QTLs for molecular breeding.

Dr Yu developed the world's first integrated genetic, physical and transcript maps of cultivated tetraploid cotton chromosomes with large insert DNA clones, molecular markers, and EST genes; and high-density cotton genetic maps of portable SSR and SNP markers. He has freely filled numerous requests from the global cotton research community for genomic resources and information. Dr Yu has received several national and international awards and published more than 100 peer-reviewed papers, which have 8,945 citations. Dr Yu has also delivered more than 60 talks/seminars to the global scientific community.

Dr Yu demonstrated dynamic, outstanding organisational capacity in a leadership role to benefit the global cotton community. This includes his election in 2005 and re-election in 2015 by the global cotton community as Chairman of the International Cotton Genome Initiative (ICGI), the first international organisation to facilitate global collaborative research work on cotton genomics and genetics.

Dr Yu led international efforts to develop and release genome sequences for G. arboreum, G. raimondii, and G. hirsutum, among other cotton species. The research opens a new paradigm in cotton genomics that revolutionises genetic improvement of cotton plants through better exploitation of genetic variation otherwise buried in Gossypium germplasm. His discovery of gene-rich islands and sub-genomic roles in upland cotton provides critical information for genetic applications. His development of new concepts and methodologies for standardising characterisation of Gossypium germplasm makes it possible for cotton researchers worldwide to collaborate, relate and utilise genetic diversity data among cotton germplasm collections.

The ICAC Recorder

Mini-Review

Global Status of Cotton Genomics and Utilisation in Improving Trait Value

Dharminder Pathak, Department of Plant Breeding and Genetics, **Yadhu Suneja**, Department of Biochemistry and **Amandeep Kaur Gill**, School of Agricultural Biotechnology
Punjab Agricultural University, Ludhiana – 141004, Punjab, India

Cotton belonging to the genus *Gossypium* is one of the most important commercial crops of the world and contributes immensely to the economy of many countries. Lint, the major product of cotton, is the chief raw material of the textile industry. In addition to the lint, edible oil and proteins in the cotton seed enhance the value of cotton as a crop. During 2018/19, cotton was planted in an area of 32.65 million hectares (ha) globally with India, China, USA, Pakistan, Brazil, and Uzbekistan accounting for 4/5 of the world cotton production of 25.694 million metric tonnes (www.icac.org). The term 'cotton' represents four cultivated *Gossypium* species *viz*., *G*. *hirsutum*, *G*. *barbadense*, *G*. *arboreum*, and *G*. *herbaceum*. Among the four species, *G*. *hirsutum*, also known as Upland/American cotton, is the dominant cotton species accounting for about 95% of the total cotton production (Trapero *et al.,* 2016). The extra-long staple cotton species, *G*. *barbadense*, is cultivated in in Egypt, USA, Sudan, Uzbekistan, China, India, Israel, Turkmenistan, Peru, etc., which accounts for 2% to 3% of the global cotton area. The diploid cultivated cotton species namely *G*. *arboreum* and *G*. *herbaceum* (collectively called Asiatic cottons) possessing short and coarse fibres, is primarily cultivated in India, Pakistan, China, Iran, Bangladesh, Myanmar, Thailand etc., in less than 1% of global cotton area.

The genus *Gossypium* includes about 50 species and are distributed in arid to semi-arid regions of the tropics and subtropics (Wendel and Cronn, 2003). The diploid *Gossypium* species (2n = 2x = 26) belong to any one of the eight (A, B, C, D, E, F, G, and K) genomes. Seven allotetraploid cotton species (2n = 4x = 52) with AD genomes have been recognised. Of these, two species, namely *G*. *ekmanianum* (Krapovickas and Seijo, 2008) and *G*. *stephensii* (Gallagher *et al.,* 2017), have been described relatively recently. A recent review on the designations of individual *Gossypium* genomes and chromosomes (Wang K *et al.,* 2018) is available. The allopolyploid cotton species originated about 1 - 2 million years ago through the hybridisation between A-genome diploid species from the old world (resembling extant *G*. *herbaceum*) and a D-genome cotton species from the new world (resembling extant *G*. *raimondii*) followed by chromosome doubling -have a monophylectic origin (Grover *et al*., 2012). Greater than threefold variation among the genome sizes of the diploid cotton species (885 MB in D-genome species through 2570 MB in K-genome species) has been observed (Hendrix and Stewart, 2005). Depending on the ease of development of inter-specific hybrids and the frequency of chromosomal recombination, *Gossypium* species have been divided into primary, secondary, and tertiary gene pools. All the tetraploid *Gossypium* species constitute the primary gene pool. The secondary gene pool is comprised of species with A, D, B, and F genomes. C, E, G, and K genome species are included in the tertiary gene pool. These are difficult to hybridise with tetraploid cotton and a low level of genetic recombination has been observed.

DNA-based markers such as restriction fragment length polymorphisms (RFLPs), random amplified polymorphic DNAs (RAPDs), amplified fragment length polymorphisms (AFLPs), simple sequence repeats (SSRs), and single nucleotide polymorphisms (SNPs) have been extensively used in cotton for various purposes including the development of genetic maps, tagging of economically important genes/quantitative trait loci (QTLs), marker aided selection (MAS), germplasm characterisation, varietal fingerprinting, etc. Construction of high-density linkage maps is one of the most important applications of molecular markers. The first molecular linkage map in cotton was reported 25 years ago by Reinisch *et al.*, (1994). Since then, a large number of molecular linkage maps using interspecific (*G*. *hirsutum* x *G*. *barbadense*, *G*. *hirsutum* x *G*. *darwinii*, *G. hirsutum* x *G. tomentosum*, *G. hirsutum* x *G. mustelinum*, *G. arboreum* x *G. herbaceum*, *G. thurberi* and *G. trilobum*, *G. davidsonii* x *G. klotzschianum*) as well as intraspecific populations (within *G*. *hirsutum*, *G. barbadense*, *G. arboreum*, *G. sturtianum*) have been generated. Selected cotton linkage maps developed since 2011 and their salient features have been presented in Table 1. The target traits reported in these maps include fibre length, fibre strength, micronaire, fibre elongation, fibre uniformity, earliness, plant height, number of monopods and sympods, number of bolls per plant, boll weight/size, lint percentage, seed index, salt and drought tolerance, photoperiod insensitivity, seed nutrients, somatic embryogenesis etc. A comprehensive review on cotton molecular breeding by Fang (2015) that includes marker development, construction of molecular linkage

Table 1: List of cotton molecular linkage maps developed since 2011

Sr. No.	*Gossypium* species involved	Parentage	Mapping population	Population size	Marker system employed	No. of linkage	No. of loci mapped	Map distance	Reference
1	*G. hirsutum*	Zhongmiansuo-12 × J8891	RIL	171	cDNA-AFLP	26	302	2477.06	Liu *et al.,* (2011)
2	*G. sturtianum*	Gos-5131 × Gos-5050	F_2	114	AFLP	20	291	1697	Lopez-Lavalle *et al.,* (2011)
3	*G. hirsutum*	FH-901 × RH-510	F_2	136	SSR	12	35	273.1	Saeed *et al.,* (2011)
4	*G. hirsutum* and *G. barbadense*	Emian22/3-79/ Emian22	BC_1	141	SSR	26	2316	4418.9	Yu *et al.,* (2011)
5	*G. hirsutum and G. barbadense*	Consensus map	-	-	SSR, AFLP, RFLP, SNP	26	8254	4070	Blenda *et al.,* (2012)
6	*G. hirsutum*	Acala Maxxa × TX2094	$F_{2:3}$	174	SNP	38	367	1688	Byers *et al.,* (2012)
7	*G. hirsutum* and *G. barbadense*	TM-1 × 3-79	RIL	186	SSR, SNP	26	2527	3430	Fang and Yu (2012)
8	*G. hirsutum*	HS46 x MARKCBUCAG8US-1-88	F_2	376	SSR	30	388	1946.22	Liu *et al.,* (2012)
9	*G. hirsutum*	MCU-5 × Siokra	F_2	244	SSR, AFLP	18	151	1189	Lopez-Lavalle *et al.,* (2012)
10	*G. hirsutum*	0-153 × sGK9708	F_2	250	SSR	26	155	959.4	Sun *et al.,* (2012)
11	*G. hirsutum*	0-153 × sGK9708	RIL	196	SSR	34	190	700.9	Sun *et al.,* (2012)
12	*G. hirsutum* and *G. barbadense*	TM-1 × 3-79	RIL	186	SSR, SNP	26	2072	3380	Yu *et al.,* (2012)
13	*G. hirsutum*	Yumian 1/CRI 35//Yumian 1/7,235	Composite cross population	180	SSR	69	978	4184.4	Zhang *et al.,* (2012)
14	*G. hirsutum* and *G. barbadense*	TM-1/ Hai7124/ TM-1	BC_1	138	AFLP, SSR, SNP, InDels, REMAP, RT, SRAP	26	3414	3667.62	Zhao *et al.,* (2012)
15	*G . hirsutum*	GX1135 x GX100-2	F_2	173	SSR	49	421	3814.3	Liang *et al.,* (2013)
16	*G. arboreum*	Ravi × Entry-17	F_2	180	SSR, RAPD	10	45	346	Shaheen *et al.,* (2013)
17	*G. barbadense*	Hai7124 × 3-79	F_2	124	SSR, SRAP, SNP	52	337	2108	Wang *et al.,* (2013)
18	*G. hirsutum* and *G. barbadense*	SG 747/Giza 75/SG 747	BIL	146	SSR	29	392	2895	Yu *et al.,* (2013)
19	*G. hirsutum* and *G. barbadense*	TM-1 × Introgressed line NM24016	RIL	98	SSR, SNP	117	841	2061	Gore *et al.,* (2014)
20	*G. hirsutum*	Yumian-1 x AcalaMaxxa	F_2	180	SSR	63	323	1617.2	Shao *et al.,* (2014)
21	*G. hirsutum*	Yumian-1 x CA3084	F_2	180	SSR	57	302	1639.9	Shao *et al.,* (2014)
22	*G. hirsutum*	Yumian-1 x TAM94L-25	F_2	180	SSR	49	262	1441.4	Shao *et al.,* (2014)
23	*G. hirsutum*	Baimian 1 x TM-1	$F_{2:3}$		SSR	37	144	1273	Wang *et al.,* (2014)
24	*G. hirsutum*	Acala-Prema x 86-1	RIL	180	SSR	54	279	1576.25	Zhiyuan *et al.,* (2014)
25	*G. arboreum* and *G. herbaceum*	SMA-4 x A-97	F_2	94	SSR, AFLP, TRAP	37	606	1109	Badigannavar and Myers (2015)
26	*G . hirsutum* and *G . darwinii*	CCRI 12-4 x (AD) 5-7	F_2	188	SSR	26	2763	4176.7	Chen *et al.,* (2015)
27	*G. barbadense* and *G. hirsutum*	3-79 x TM-1	F_2	118	SNP	26	19191	4439.6	Hulse-Kemp *et al.,* (2015)
28	*G. hirsutum*	Phytogen 72 x Stoneville 474	F_2	93	SNP	26	7171	3499	Hulse-Kemp *et al.,* (2015)
29	*G. hirsutum*	Yumian 1 x T586	RIL	270	SSR	26	1675	3338.2	Liu *et al.,* (2015)
30	*G. hirsutum*	GX1135 x GX100-2	RIL	177	SSR	14	581	3714.4	Shang *et al.,* (2015)
31	*G. hirsutum*	CCRI-35 x Yumian-1	RIL	180	SSR	26	1274	3076.4	Tan *et al.,* (2015)
32	*G. hirsutum*	Yumian-1 x 7235	RIL	180	SSR	26	1582	2842.06	Tang *et al.,* (2015)
33	*G. hirsutum*	DH962 X Ji Mian 5	F_2		SNP, SSR, InDels	50	1013	3004.71	Wang *et al.,* (2015b)
34	*G . hirsutum*	DH962 x Ji Mian 5	RIL	178	SSR, InDels, SNP	59	616	2016.44	Wang *et al.,* (2015a)
35	*G. hirsutum*	Acala Prema x 86–1	RIL	161	SNP	103	4153	3500	Wang Y *et al.,* (2015)
36	*G. hirsutum*	W10 x TM-1	F_2	140	SSR	42	411	2300.41	Xu *et al.,* (2015)
37	*G. hirsutum* and *G. barbadense*	CCRI36 × G2005	RIL	137	SNP, SSR	26	6434	4071.98	Jia *et al.,* (2016)
38	*G. hirsutum* and *G. tomentosum*	CRI12-2 x P0601211	F_2	188	SSR	26	3093	4365.3	Khan *et al.,* (2016)
39	*G. hirsutum* and *G. barbadense*	Emian22/ 3–79/Emian22	BC_1	141	SSR	26	5152	4696.03	Li X *et al.,* (2016)
40	*G. hirsutum*	HS46 × MARCABUCAG8US-1-88	RIL	188	SNP	26	2618	1784.28	Li C *et al.,* (2016)

Table 1: List of cotton molecular linkage maps developed since 2011

Sr. No.	*Gossypium* species involved	Parentage	Mapping population	Population size	Marker system employed	No. of linkage	No. of loci mapped	Map distance	Reference
41	*G. hirsutum*	0-153 × sGK9708	RIL	196	SSR	76	793	4110	Jamshed *et al.*, (2016)
42	*G. hirsutum*	GX1135 × GX100-2	RIL	180	SSR	39	308	3048.4	Shang *et al.*, (2016)
43	*G. hirsutum*	GX1135 × GX100-2	RIL	177	SSR	32	622	3889.9	Shang *et al.*, (2016)
44	*G. hirsutum*	TM-1 x (TX-0256 + TX-1046)	CSIL	115	SSR	37	713	3594.70	Zhang *et al.*, (2016)
45	*G. hirsutum* and *G. mustelinum*	PD 94042 × (AD)4-08	F_2	92	SSR, STS, CAPS, DNA probes	26	1055	5595	Wang B *et al.*, (2016)
46	*G. hirsutum* and *G. barbadense*	Fibre Max 966 × CIR1348	F_2	174	SSR	29	262	4294	Gomez *et al.*, (2016)
47	*G. hirsutum*	LU28 × ZHONG213	F_2	170	SNP	26	3978	2480	Li L *et al.*, (2017)
48	*G. hirsutum*	CCRI35 × Nan Dan	$F_{2:3}$	277	SNP	26	5178	4768.098	Diouf *et al.*, (2017)
49	*G. hirsutum*	Phytogen 72 x Stoneville 474	RIL	132	SNP	26	7059	3966	Ulloa *et al.*, (2017)
50	*G. hirsutum*	Stoneville 474 x Phytogen 72	RIL	104	SNP	26	6320	3862	Ulloa *et al.*, (2017)
51	*G. hirsutum*	Phytogen 72 x Stoneville 474	F_2	93	SNP	26	7034	3597	Ulloa *et al.*, (2017)
52	*G. hirsutum*	0-153 x sGK9708	RIL	196	SNP, SSR	26	2393	2865.73	Zhang *et al.*, (2017)
53	*G . darwinii*	Wild type *G. darwinii* accession x photo-insensitive mutant counterpart	F_2	129	SSR	25	194	1158.5	Kushanov *et al.*, (2017)
54	*G. hirsutum*	Z571 × CCR149	F_2	188	SNP	26	3187	3828.55	Qi *et al.*, (2017)
55	*G. hirsutum* and *G. barbadense*	Pima S-7 × 5917	RIL	143	SNP	26	3557	3076.23	Fan *et al.*, (2018)
56	*G. thurberi* and *G. trilobum*	*G. thurberi* x *G. trilobum*	F_2	188	SSR	13	849	1012.458	Li S *et al.*, (2018)
57	*G. davidsonii* and *G. klotzschianum*	D3K-57 × D3D-1	$F_{2:3}$	188	SSR	13	728	1480.23	Kirungu *et al.*, (2018)
58	*G. hirsutum* and *G. tomentosum*	CRI-12/AD3-00/CRI-12	BC_2F_2	200	SNP	26	10888	4191.3	Magwanga *et al.*, (2018)
59	*G. hirsutum*	Yumian 1 × Acala Maxxa	RIL	180	SNP	26	12116	3741.81	Tan *et al.*, (2018)
60	*G. hirsutum*	Lumianyan 28 × Xinluzao24	RIL	231	SSR, SNP	26	4851	2477.99	Liu *et al.*, (2018)
61	*G. barbadense*	Suvin × BCS 23-18-7	F_2	185	SNP	21	460	1219.4	Kumar *et al.*, (2019)
62	*G. hirsutum*	JKC737 × JKC725	RIL	180	SSR	18	120	2883.8	Ravichandran *et al.*, (2019)
63	*G. hirsutum*	CRI50 × CRI60	F_2	198	SNP	26	5280	3848.29	Zhang *et al.*, (2019)
64	*G. hirsutum* and *G. barbadense*	DS28 × SBYF425	RIL	136	SNP	26	1430	3149.8	Ramesh *et al.*, (2019)
65	*G. hirsutum* and *G. barbadense*	CRI36/Hai7124/CRI36	BIL	250	SNP	26	7709	3433.24	Ma *et al.*, (2019)

maps, mapping of economically important traits such as fibre quality, yield, disease resistance, seed traits/constituents, plant morphology, male-sterility etc. and marker assisted selection is available. Here, we give an update on the mapping of genes/QTL linked with male-sterility, root knot nematode, and verticillium wilt (Table 2).

The ultimate physical map of an organism is the nucleotide sequence of its genome. It provides an insight on the ultra-structure of the genome, its evolution, etc. Moreover, no gene can escape identification. Genomes of a few *Gossypium* species have been sequenced. *G. raimondii* (DD) was the first cotton (progenitor) species to have been sequenced (Paterson *et al.,* 2012; Wang *et al.,* 2012). This was followed by the genome sequencing of *G. arboreum* (AA) (Li *et al.,* 2014). The genome sequence of the cultivated allotetraploid cotton species namely *G. hirsutum* (AD)1 and *G. barbadense* (AD)2 became available in 2015 (Li *et al.,* 2015; Zhang T *et al.*, 2015; Liu X *et al.,* 2015). However, there are differences in the sequenced and assembled draft genomes of the cotton species with respect to chromosome length and annotated genes that needs consideration. Availability of genome sequences of cotton species has facilitated the development of a 63K (Hulse-Kemp *et al.*, 2015) and an 80K single nucleotide polymorphism (SNP) array (Cai *et al.,* 2017). These SNP chips have been used for the construction of highly dense cotton linkage maps, mapping of fibre quality, yield and component traits, male-sterility, disease resistance, lint percentage, salt stress etc. (Table 3). Resequencing of more than 800 cotton accessions has also been reported (Page *et al.,* 2016; Fang *et al.,* 2017a, 2017b; Wang M *et al.*, 2017).

Genetic variation serves as the raw material and is central to plant-breeding activities. Crop scientists around the globe are working to identify novel allelic variants in the existing germplasm (worldwide collection of cultivars, landraces, and un-adapted/wild accessions) or create new sources of genetic variation by either using traditional breeding approaches (intensive selection and hybridisation), de-

Table 2: Mapping of root knot nematode, verticilliam wilt and male-sterility in cotton

Gossypium species involved	Parentage	Population type	No. of individuals	Marker system employed	Reference	Remarks
Root Knot Nematode (RKN)						
G. hirsutum and *G. barbadense*	Fibre Max 966 × CIR1348	F_2	174	SSR	Gomez *et al.,* (2016)	Four chromosomal regions harboring QTLs for RKN resistance in a novel *G. barbadense* source were identified.
Verticillium Wilt (VW)						
G. hirsutum and *G. barbadense*	SG 747/Giza 75/SG727	BIL	146	SSR	Zhang *et al.,* (2015)	Ten VW resistance QTLs on eight chromosomes were identified. These QTLs have been reported in previous studies also.
G. barbadense and *G. hirsutum*	Hai1/CCRI36/CCRI36	BC_1F_1, BC_1S_1, BC_2F_1	135	SSR	Shi *et al.,* (2016)	Forty eight QTLs associated with VW on 19 chromosomes were detected. 37 QTLs were positive and 6 were observed to be stable. The study validated six previously identified QTLs.
G. hirsutum	-	Germplasm Accessions	318	SNP	Fang *et al.,* (2017b)	Three loci for VW resistance were detected in the D-subgenome. A candidate gene (*Gh_D06G0687*) encoding an NB-ARC domain-containing disease resistance protein was identified.
G. hirsutum	-	Germplasm Accessions	299	SNP	Li T *et al.,* (2017)	17 Trait-SNP associations including one candidate gene for VW resistance were detected.
G. hirsutum	-	Germplasm Accessions	120	SNP	Zhao Y *et al.,* (2017)	Three QTLs and several minor loci for VW resistance were selected.
G. hirsutum	0-153 x sGK9708	RIL	196	SNP	Palanga *et al.,* (2017)	119 QTLs of disease index and of disease incidence were detected on al the chromosomes except 13.
G. hirsutum	Introgressed line SuVR043 × Sumian 8	F_2	176	SSR	Zhao *et al.,* (2018)	Two VW resistance QTLs were detected.
Male-sterility						
G. hirsutum	DPGh98018 x DPGh04651	Sibcross F_2, F_3 populations derived from 14 sibcross F_2 populations	For fine mapping, population of 1044 inbred lines was used.	SNP	Feng *et al.,* (2015)	Four SNPs for both *ms5* and *ms6* gene regions were identified and validated.
G. hirsutum and G. barbadense	Kang A x 601588	F_2	180	SSR	Wang P *et al.,* (2016)	*ms14* conditioning male sterility mapped on chromosome 2
G. arboreum	Ga TGMS-3	Fertile and sterile plants of Ga TGMS-3	-	SSR	Sekhar *et al.,* (2016)	NAU2176, NAU2096 and BNL1227 markers linked to TGMS in *G.arboreum* line Ga TGMS-3 were identified.
G. hirsutum	RC403 x RC370	BC_1	4 bulks	Cotton63k SNP	Raja *et al.,* (2017)	Four SNP markers found to be linked with male sterility
G. hirsutum	CCRI 9106 × Letu 603	F_2	186	SSR	Zhang M *et al.,* (2017)	*ys-1* gene governing photoperiod-sensitive male sterility was identified on chromosome D12 within markers NAU3442 and CGR6339.
G. hirsutum	CMS-2074A, CMS-2074S and restorer lines	Mioochomdrial genomes	-	mt RNA sequencing	Li S *et al.,* (2018)	Four novel and recombinant ORFs were found to be potential candidates for CMS.
G. hirsutum	3096A x 866R	F_2	fertile and sterile bulks	SLAF-seq	Zhao C *et al.,* (2018)	Fertility restorer candidate genes located on 1.35 Mb of chromosome D05, and 20 candidate genes were identified.

velopment of cis-genic and trans-genic plants, reverse genetics tools like T-DNA insertional mutagenesis, Targeting Induced Local Lesions In Genome (TILLING), transposon tagging etc., or by using the more recent genome editing tools. While genetic variation introduced through earlier approaches was essentially dictated by chance, genome editing ensures introduction of the variation at a specific genomic locus. Genome editing mediated by site specific nucleases (SSNs) — like meganucleases/homing endonucleases, zinc-finger nucleases (ZFNs); transcription activation like effector nucleases (TALENs); clustered, regularly interspaced, short palindromic repeats (CRISPR)/CRISPR associated (Cas) 9 — allows targeted genome modifications. Though the feasibility of use of these SSNs for site specific cleavage was first demonstrated in early 1990s, one of the major bottlenecks in their successful implementation was limited information available with respect to the genomic landscape, functional characterisation of gene or specific knowledge on relevance of the particular gene/network in influencing the trait of interest. However, with the availability of whole genome sequence for most of the crops (alongside functional annotation for model plant species like *Arabidopsis* and rice), and ease of conducting *de novo* transcriptome analysis, it is now possible to identify target genomic loci and accordingly hit the DNA at the core nucleotide level. The post-genomic era is truly seeing the upsurge of genome-editing-based precision breeding. A brief description concerning RNAi, TILLING and different site-specific nucleases is given below.

Table 3: Reports on the development and use of cotton SNP chips							
Sr. No.	***Gossypium* species involved**	**Parentage**	**Population type**	**No. of individuals**	**No. of loci**	**Reference**	**Remarks**
CottonSNP63K							
1	*G. barbadense* and *G. hirsutum*	3-79 x TM-1	F_2	118	19191	Hulse-Kemp *et al.,* (2015)	Development of highly dense intraspecific and interspecific maps
	G. hirsutum	Phytogen 72 x Stoneville 474	F_2	93	7171		
2	*G. hirsutum*	HS46 x MARCABUCAG 8US-1-88	RIL	188	2618	Li C *et al.,* (2016)	Detection of 71 fibre quality and yield traits QTLs
3	*G. hirsutum*	-	Germplasm accessions	85	-	Zhu *et al.,* (2016)	Narrowing down of okra leaf locus
4	*G. hirsutum*	-	Germplasm accessions	201	23254	Handi *et al.,* (2017)	Identification of markers associated with yield, yield components and fibre quality traits
5	*G. hirsutum, G. arboreum, G. amourianum, G. longicalyx, G. raimondii, G. thurberi, G. trilobum, G. barbadense, G. ekmanianum, G. mustelinum*and *G. tomentosum*	-	Germplasm accessions	390	33507	Hinze *et al., (*2017)	Assessment of diversity and identification of loci linked to seed protein content
6	*G. hirsutum*	-	Germplasm accessions	503	11975	Huang *et al., (*2017)	324 SNPs and 160 QTL regions identified as associated with 16 agronomic traits
7	*G. hirsutum*	RC403 x RC370	BC_1	4 bulks	2855	Raja *et al.,* (2017)	Four SNP markers found to be linked with male sterility
8	*G. hirsutum*	-	Germplasm accessions	719	10511	Sun *et al.,* (2017)	Detection of 46 significant SNPs associated with fibre quality traits
9	*G. hirsutum*	Phytogen 72 x Stoneville 474	F_2, RIL, Reciprocal RIL	93 132 104 respectively	7034 7059 6320	Ulloa *et al.,* (2017)	Generation of consensus map and information on chromosome arrangement of genic and non-genic SNPs
10	*G. hirsutum*	sGK9708x 0-153	RIL	196	2393	Zhang *et al.,* (2017)	Identification of fibre strength QTLs
11	*G. hirsutum*	-	Germplasm accessions	120	18,726	Zhao *et al.,* (2017)	Detection of three verticillium wilt resistance QTLs
12	*G. hirsutum*	-	Germplasm accessions	93	15369	Ma *et al.,* (2018)	Ten SNPs forming five QTL regions showing correlations with fibre-length growth rate
13	*G. barbadense* and *G. hirsutum*	Pima S-7 × Sicala 40	NILs	2 bulks	5426	Zhu *et al.,* (2018)	Five genetic loci, including a major contributing locus containing MYB25-like_Dt associated with Gb fuzzless seeds
14	*G. barbadense*	Suvin × BCS 23-18-7	F_2	185	460	Kumar *et al.,* (2019)	Identification of yield components and fibre quality QTLs
15	*G.hirsutum* and *G. barbadense*	DS28 × SBYF425	RIL	178	1430	Ramesh *et al.,* (2019)	QTL detected for fibre quality and yield components
16	*G. hirsutum*	-	Germplasm accessions	276	10660	Song *et al.,* (2019)	QTLs and two candidate genes identified for lint percentage

Table 3: Reports on the development and use of cotton SNP chips							
Sr. No.	*Gossypium* species involved	Parentage	Population type	No. of individuals	No. of loci	Reference	Remarks
CottonSNP80K							
17	*G. hirsutum* and *G. barbadense*	-	Germplasm accessions	332	59502	Cai *et al.,* (2017)	Phylogenetic analysis and SNPs associated with salt stress identified
18	*G. hirsutum*	-	Germplasm accessions	169	49650	Li *et al.,* (2018a)	Two potential candidate genes for early maturity identified
19	*G. hirsutum*	-	Germplasm accessions	169	53848	Li *et al.,* (2018b)	Identification of fibre quality QTLs
20	*G. hirsutum*	Xinluzao24 x Lumianyan28	RIL	231	4729	Liu *et al.,* (2018)	QTLs for fibre quality and yield components trait identified
21	*G. hirsutum*	Yumian 1 × AcalaMaxxa	RIL	180	12116	Tan *et al.,* (2018)	Identification of fibre quality QTL

RNAi (RNA interference)

RNAi was first identified as a biologically conserved response to pathogenic attack from studies in *C. elegans* (Fire *et al.,* 1998). In response to parasitic/pathogenic invasions, dsRNA is cleaved by Dicer enzyme into small 21-24 nucleotide long RNAs (siRNA, miRNA etc.) which when loaded on to RNA Induced Silencing Complex (RISC) recognise and degrade complementary mRNA preventing damage to the plant. This innate mechanism of post transcriptional gene silencing and potential of an artificial dsRNA to silence a gene (nucleotide sequence known) has been extensively exploited by scientists as a reverse genetics approach to reduce the expression of selected genes with the aim of producing the desired phenotype. Abdurakhmonov and colleagues (2016) have provided a comprehensive review on the use of RNAi for functional genomics purposes and practical cotton improvement in terms of enhancing disease resistance, abiotic stress tolerance, fertility and embryogenesis, quality fibre development and enhancing seed and oil quality. RNAi was employed to enhance resistance of cotton plants to the cotton bollworm *Helicoverpa armigera* by expressing insect-derived cytochrome P450 monooxygenase gene (dsCYP6AE14) (Mao *et al.,* 2011). Further, to enhance nutritional quality of cottonseed oil, endogenous genes like FAD2 and SAD1 were functionally disrupted to increase the content of oleic acid and stearic acid, respectively (Liu *et al.,* 2002). δ-cadinene synthase which catalyses the first step in the formation of cadenine was silenced by RNAi strategy to reduce the content of gossypol in cottonseed (Sunilkumar *et al.,* 2006). Recently, a genetically engineered event (TAM66274) for ultra-low gossypol levels in the cottonseed has been approved in the United States (c.f. Hagenbucher *et al.,* 2019).

TILLING (Targeting Induced Local Lesions in Genomes)

Soon after the ground-breaking discovery of RNAi by Fire and Mello in 1998, a report demonstrating successful use of chemical mutagenesis to characterise chromo-methylase gene in Arabidopsis laid the foundation of another very effective and a high-throughput reverse genetics approach now popularly known as Targeting Induced Local Lesions in Genome (TILLING) (McCallum *et al.,* 2000). Treating seeds with ethyl-methane-sulphonate (EMS), a well-known chemical mutagen, leads to point mutations particularly transitions (G/C : A/T) which result in the creation of bi-allelic mutations in the form of SNPs (single nucleotide polymorphisms). The genetic variation so induced may be evaluated in the M_2 population by amplifying the target gene fragments from pooled DNA samples and observing the cleavage of heteroduplexes so formed by cel1 nuclease (recognises the mismatch and cleaves the DNA on the 3'end of the mismatch). This shall allow the identification of new allelic variants and their precise location in the genome where the mutation has been introduced which may then be confirmed by sequencing. Aslam *et al.,* (2016) developed TILLING population of two cotton cultivars PB-899 and PB-900 in *G. hirsutum* using different concentrations of EMS (0.3 and 0.2% respectively). M_1 and M_2 populations were phenotyped for several important traits like branching pattern, leaf morphology, disease resistance (particularly CLCuD), photosynthetic lesions and flower sterility. When screened at the molecular level, highest mutation frequency was found in NAC transcription factor gene. This cotton TILLING platform (COTIP) has created wealth of mutants to be screened and used for improving traits of agronomic significance in cotton. Though TILLING is a relatively simple, non-transgenic approach to create new sources of genetic variation, the method is strongly limited by the need to phenotype thousands of plants in the population to identify the desired phenotype. A key defining feature of TILLING is an element of randomness and chance mutation at many locations in the genome, rather than a specific gene sequence that may be targeted using this strategy.

MegaN (Meganucleases/Homing Endonucleases)

Among the various SSNs, meganucleases/homing endonucleases were the first enzymes to be used for genome editing in plants (Smith *et al.,* 2006). They belong to the family of naturally occurring restriction endonucleases that can recognise and cut larger unique DNA sequences (from 12-40 bp) in most genomes (Paques and Duchateau 2007). Four different families of meganucleases have been identified- LAGLIDADG, His-Cys box, GIY-YIG, and the HNH family. Commonly used homing endonucleases- I-CreI and I-SceI belong to LAGLIDADG family. Overlapping nature of DNA binding and nuclease domain in these meganucleases poses difficulty in engineering these proteins for targeting a specific site in the genome as target site specificity is generally achieved at the cost of reduced catalytic activity (Paques and Duchateau 2007). Further, there is need to incorporate recognition sites of these endonucleases in the plant genome as these do not occur naturally in the plant genome, which is the major limitation in the use of these endonucleases. Meganucleases have been modified for use in genome editing by different companies such as Cellectis and Precision Biosciences Inc. In cotton, meganucleases were employed to edit *hppd* and *epsps* genes to enhance herbicide resistance (D'Halluin *et al.,* 2013). Since each new GE target requires an initial protein engineering step to produce a custom nuclease, meganucleases have proved technically challenging to work with and have been hindered by patent disputes (Smith *et al.,* 2011).

ZFNs (Zinc-finger nucleases)

The second category of sequence-specific nucleases includes synthetic hybrid protein, namely zinc-finger nucleases (ZFNs) and transcription activator like effector nucleases (TALENs). Zinc fingers, an important class of transcription factors found in the eukaryotes are known to recognise specific DNA sequence elements (each zinc finger recognises three consecutive nucleotides) and promote gene transcription. Made up of a 30 amino acid long Cys_2-His_2 motif (Beerli and Barbas 2002), each zinc finger unit consists of 3-4 modules that may be tailor-joined to detect a specific DNA sequence in the genome. Thus, each zinc finger nuclease (ZFN) contains an engineered array of zinc finger motifs that serve as DNA recognition and binding domains (Bibikova *et al.,* 2003, Carroll 2011) attached via conserved linker sequences to a catalytic domain of monomeric *Fok1* endonuclease. The first report on successful demonstration of cleavage activity of ZFNs was published in 1996 (Kim *et al.,* 1996). Since then, several methods like modular assembly, Oligomerised Pooled Engineering (OPEN) and Context Dependent Assembly (CoDA) have been developed for linking together individual zinc fingers and custom synthesising ZFNs. Since *Fok1* functions as a dimer, a pair of zinc finger arrays (ZFAs) is designed to bind to target sequence (with a spacer sequence of 5-7 nucleotides between them) to create a double strand break (DSB). For many years, ZF protein technology was the only approach available to create custom site-specific DNA binding protein and enzymes. Despite its unique ability to recognise a specific site and cleave DNA segment, the method suffered from several limitations like identification of a unique triplet nucleotide by each ZF, content dependent binding of ZFN to the target site, difficulty in designing and synthesising protein-based nucleases for each individual site and predicted cleavage of off-target sites in complex plant genomes. To our knowledge, there is no report of use of ZF nucleases for genome editing in cotton.

TALENs (Transcription-activator like effector nuclease)

The next wave of innovation in the field of protein-engineered genome editing tools was the use of TALE proteins — transcription activator like effector proteins to design synthetic nucleases. The landmark discovery that resulted in the identification of DNA recognition code for establishing target DNA specificity in TALEs paved the way for combining DNA-protein interaction ability of TALE and catalytic activity of *Fok1* nuclease to design and synthesise next generation sequence specific nucleases i.e. TALENs (Christian *et al.,* 2010). TALE proteins are virulence factors released by *Xanthomonas* bacteria that bind to host DNA and alter gene transcription in plants. The 'TALE' component that serves as the DNA binding domain of TALENs is made up of N- terminal secretion (translocation signal), C-terminal nuclear localisation signal (NLS), an acidic transcription activation domain (AD) and a central DNA binding domain (DBD). Within this central protein domain reside up to 30 tandem arrays of highly conserved 33-35 amino acids long repeats. The unique combination of amino acids at position 12 (P-12) and position 13 (P-13) within this repeat is responsible for the recognition of a specific nucleotide i.e. repeat variable di residues (RVDRs) like HD = C, NG = T, NI = A, NN, NK and NH = G (Boch *et al.,* 2009, Moscou and Bgdanove 2009). Keeping this RVDR code as the base, order of stitching TALE molecules may be worked out to design custom binding sites. A pair of TALEs may be associated with *Fok1* dimer to recognise a contiguous stretch of DNA and mediate cleavage on both the strands. The double strand break so created may then be ligated by a cell's innate repair mechanism like homologous recombination (HDR) or non-homologous end joining (NHEJ) pathways.

Several methods have been developed to assemble TALEs. These include-ligation independent standard cloning (Huang *et al.,* 2011, Reyon *et al.,* 2012a), HH ligation based golden gate assembly method (Cermak *et al.,* 2011, Weber *et al.,* 2011) and high throughput solid phase assembly method (Reyon *et al.,* 2012b). Unlike ZFNs where a linker

connecting each zinc finger has to be re-engineered, assembling TALE repeats does not suffer from any such limitation. TALEs can easily be linked together to recognise any user defined sequence. Further, where each ZFN may recognise a triplet nucleotide, ability of TALEs to identify a single base allows greater flexibility in designing TALE arrays. However, repetitive nature of DBDs and larger size of the TALE protein poses hindrance to engineering and delivering TALENs to the target cell. Moreover, the need for the incorporation of a 'T' before the first nucleotide at the target site and presence of around 15-30 bases between two TALEN binding sites may affect its activity. Khan *et al.* (2017) have advocated the use of engineered nucleases including TALENs for the suppression of cotton leaf curl disease (CLCuD), a whitefly transmitted viral disease prevalent in Pakistan, North Indian cotton growing states, and China.

CRISPR/Cas9 system

Unlike the DNA-protein interactions that formed the basis of target site recognition in meganucleases, Zinc finger nucleases and TALENs, the most recent addition to class of site specific nucleases i.e. CRISPR/Cas9 system edits genomes by mediating the formation of a DNA-RNA hybrid and its subsequent interaction in the presence of a nuclease. Greater ease of designing and robust nature of this RNA guided endonuclease (RGENs) has allowed rapid adoption of CRISPR/Cas9. Clustered Regularly Interspaced Short Palindromic Repeats CRISPR and CRISPR-associated enzyme (Cas) is a simple three-component system made of crRNA, tracrRNA and an endonuclease protein like Cas9 to induce targeted genetic alterations (Datsenko *et al.,* 2012, Jinek *et al.,* 2012). The first evidence on the existence of CRISPR can be traced back to a Japanese discovery of the presence of a 29-nucleotide long repeat sequence in the genome of *E. coli* (Ishino *et al.,* 1987). Since then, these tandem repeat arrays as a remnant of phage attack have been found in several bacteria and archae bacteria and have been named CRISPR (Jansen *et al.,* 2002). Ability of the CRISPR/Cas system to acquire and integrate segments (spacers) of foreign DNA during the first attack and express them (in the form of CRISPR RNAs crRNA and trans-activating-tra-crRNA) along with an endonuclease to degrade exogenous DNA in the subsequent attack forms the basis of native immune system functional in these organisms (Pourcel *et al.,* 2005, Mojica *et al.,* 2005 and Barrangou *et al.,* 2007). This property of bacteria has been successfully exploited by scientists to design a 20-nucleotide gRNA (complementary to the target site), assembling into a chimeric gRNA scaffold (100-nucleotide long) and using type II Cas9 (obtained from *Streptococcus pyogenes*) for sequence-directed genome editing (Doudna and Charpentier 2014).

In sharp contrast to *Fok1* nuclease that cleaves as a dimer, the monomeric form of Cas9 is able to direct cleavage on account of two nuclease activities vested in it — HNH nuclease domain that cleaves the complementary strand and RuvC-like domain that cleaves the non-complementary strand of DNA. The double-strand break it induces may then be joined by HDR or NHEJ. However, the presence of a characteristic protospacer adjacent motif (PAM) (e.g. 5'-NGG-3' in case of Cas9 from *Streptococcus pyogenes*) downstream of the recognition site (that may interact with PAM interacting domain in Cas9 and assist in its binding to target region) is an essential requirement for cleavage (Jinek *et al.,* 2012). Once recognised, the specificity of gRNA binding is governed by 'seed sequence' present approximately 12 base pairs upstream of PAM. Mismatch between gRNA and DNA within this region may severely affect binding, formation of gRNA-DNA duplex and cleavage by Cas9. Further, higher GC content of gRNA-DNA hybrid has been found to support and stabilise binding between the two. Presence or addition of one or two 'G' at the 5' end of 20 nucleotide long gRNA ensures better on-target cleavage efficiency.

The simplicity of gRNA designing, smaller length of gRNA, amenability to multiplex several gRNAs with a single Cas9 to cleave multiple targets are some of the benefits offered by Cas9 over MegaN, ZFNs and TALENs. However, the possibility of a 20-nucleotide long gRNA to bind at a site other than the target site (in large complex genomes) leading to off-target mutations has emerged as one of the major limitations of CRISPR/Cas9 strategy. Designing gRNA thus becomes very critical in successful implementation of this genome editing tool. Paired gRNA approach (i.e. using 2 gRNA to target a specific region) and replacement of Cas9 with Cas9 nickase (that cleaves just one and not both the strands of DNA) has been found to lower down the rates of off-target mutations.

The use of CRISPR/Cas9 for genome editing in plants was first reported in 2013 (Li *et al.*, 2013; Shan *et al.,* 2013). The use of this technology in cotton is still in its early stages and few reports demonstrating the potential use of CRISPR/Cas9 for genome editing in cotton are available (Table 4). Majority of these studies essentially represent a proof-of-concept reporting feasibility of targeting a polyploid complex genome of cotton for editing. In most of the cases, visual marker genes like *DsRED*, green fluorescent protein (*GFP*), *phyotene desaturase* or *Cloroplastos alterados* that upon knock-out result into albino phenotypes have been edited using CRISPR/Cas9. The possibility of editing *GhMYB25*-like gene involved in pollen development was explored in cotton (Li C *et al.,* 2017). Wang Y *et al.,* (2017) used CRISPR/Cas9 to knock-out *GhARG* involved in the catabolism of amino acid arginine. Maintenance of appreciable amount of arginine allowed greater synthesis of nitric oxide on account of NOS activity. This allowed improved lateral root growth in cotton plants. Likewise, *GhALARP* concerned with production of alanine rich protein allowed the effect of its knock-out to be stud-

Table 4: Reports of CRISPR/Cas9 on genome editing in cotton

Sr. No.	Cotton species	Target gene	Trait	Delivery method	Knock-out/ HR	Remarks	off-target detection	Reference
1	*G. hirsutum*	*GFP*	marker gene for identification of transgenics	Vector based Agrobacterium mediated transformation	Knock-out generated	Of the nine T_0 plants, two plantlets showed homozygous, while rest revealed bi-allelic indels	off-target detection not carried out	Janga *et al.,* (2017)
2	*G. hirsutum*	*GhCLA1*	Development of chloroplast	Vector based Agrobacterium	-	With insertions and deletions detected, mutation efficiency varied from 47.6-81.8% in edited plants	No off-target edits detected in	Chen *et al.,* (2017)
		GhVP	vacuolar proton pump concerned with regulating cell expansion; H+ electrochemical gradient and secondary active transport of ions	mediated transformation			30 of the 64 loci examined	
3	*G. hirsutum*	*GhPDS*	Biosynthesis of chlorophyll	Vector based Agrobacterium	-	Simultaneous editing of homeologous genes in cotton; 50-61 nucleotide long gene fragment deletion in *GhPDS* locus using paired gRNA approach; multiple gene edits created simultaneously (*GhEF-1* and *GhPDS*)	off-target detection not carried out	Gao *et al.,* (2017)
		GhCLA1	Development of chloroplast	mediated transformation				
		GhEF1	Elongation factor-1 involved in protein synthesis					
4	*G. hirsutum*	*GhMYB25*	transcription factor involved in pollen development	Vector based Agrobacterium mediated transformation	-	14.2-21.4% truncation events observed	No off-target edits detected in two putative off-target sequences	Li C *et al.,* (2017)
5	*G. hirsutum*	*GhARG*	Arginase involved in the catabolism of arginine, resulting in its decreased concentration which reduces the activity of NOS enzymes increases NO concentration improving lateral root development	Vector based Agrobacterium mediated transformation	knock-outs generated	Gene editing created variants that varied from 1-44 bp deletions	off-target detection not carried out	Wang Y *et al.,* (2017)
6	*G. hirsutum*	*DsRED2*	Discosoma red fluorescent protein 2	Vector based Agrobacterium	knock-outs generated	Multiplexing based CRISPR/Ca9 mediated genome editing performed	No off-target edits detected	Wang P *et al.,* (2018)
		GhCLA1	Development of chloroplast	mediated transformation		efficiency of gene editing at target site varied from 66.7-100%		
7	*G. hirsutum*	*ALARP*	alanine rich proteininvolved in the development of cotton fibre	Vector based Agrobacterium mediated transformation	-	Upto 55 bp deletion and 99-bp insertion observed	No off-target edits detected in 15 predicted sites analyzed	Zhu S *et al.,* (2018)
8	*G. hirsutum*	*GhPDS*	Biosynthesis of chlorophyll	Vector based Agrobacterium mediated transformation	-	Regions of U6 genes used as promoter to drive the expression of sgRNA-PDS and targeting efficiency compared; *proGhU6.3.3* generated 6-7 times higher expression than *proAtU6-29*	off-target detection not carried out	Long *et al.,* (2018)

ied in development of cotton fibre. These studies have provided sufficient evidence that CRISPR/Cas9 can be effectively used for genome editing in cotton setting up the stage for using CRISPR/Cas9 strategy to enhance growth, yield, fibre and seed quality characteristics of cotton. Khan *et al.,* (2018) envisages the use of this technology to target signal transduction pathways, genes for metabolite production for developing climate resilient cotton genotypes. This RNA-guided endonuclease (RGEN) may be used for protecting plants against bacterial and viral attacks, especially by the cotton leaf curl virus (CLCuV) (Iqbal *et al.,* 2016). Several uses of CRISPR/Cas9 are being contemplated in cotton. The commercial value of cottonseed oil may be enhanced by targeting fatty acid desaturase genes using CRISPR/Cas9 to increase the content of stearic acid and oleic acid. Further, palatability of cotton seed may be enhanced by using genome editing to reduce the content of gossypol, a toxic terpenoid compound which inhibits cotton-seed utilisation as a feed for livestock despite its rich protein content. In addition to practical cotton improvement, CRISPR/Cas9 offers immense potential to undertake functional genomic studies in cotton. Epigenetic modifications that distinguish present day domesticated cotton varieties from its wild accessions may be used to fine tune the expression of specific set of genes (using CRISPRi and/or CRISPRa) to study their role in plant development and stress resilience.

References

Abdurakhmonov I Y, Ayubov M S, Ubaydullaeva K A, Buriev Z T, Shermatov S E, Ruziboev H S, Shapulatov U M et al. (2011) RNA interference for functional genomics and improvement of cotton (*Gossypium sp.*) *Front Plant Sci* 7:202. doi: 10.3389/fpls.2016.00202

Aslam U, Cheema H M N, Ahmad S, Khan I A, Malik W and Khan A A (2016) COTIP: Cotton TILLING Platform, a resource for plant improvement and reverse genetic studies. *Front Plant Sci* **7**:1863 doi: 10.3389/fpls.2016.01863

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

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero D A et al. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**:1709-12.

Beerli R R and Barbas C F 3rd (2002) Engineering polydactyl zinc-finger transcription factors. *Nat Biotechnol* **20**:135-41.

Bibikova M, Beumer K, Trautman J K and Carroll D (2003) Enhancing gene targeting with designed zinc finger nucleases. Science **300**:764.

Blenda A, Fang D D, Rami J F, Garsmeur O, Luo F and Lacape J M (2012) A high density consensus genetic map of tetraploid cotton that integrates multiple component maps through molecular marker redundancy check. *PLoS One* **7**: e45739.

Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T et al. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**:1509-12. doi: 10.1126/science.1178811

Byers R L, Harker D V, Yourstone S M, Maughan P J and Udall J A (2012) Development and mapping of SNP assays in allotetraploid cotton. *Theor Appl Genet* **124**:1201-14.

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

Carroll D (2011) Genome engineering with zinc-finger nucleases. Genetics **188**:773-82. doi: 10.1534/genetics.111.131433

Cermak T, Doyle E L, Christian M, Wang L, Zhang Y, Schmidt C, Baller J A et al. (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* **39**: e82.

Chen X G, Lu X K, Shu N, Wang S, Wang J J, Wang D L, Guo L X et al. (2017) Targeted mutagenesis in cotton (*Gossypium hirsutum* L.) using the CRISPR/Cas9 system. *Sci Rep* **7**:44304.

Chen H, Khan M K, Zhou Z, Wang X, Cai X, Ilyas M K, Wang C *et al.* (2015) A high-density SSR genetic map constructed from a F_2 population of *Gossypium hirsutum* and *Gossypium darwinii*. *Gene* **574**(2):273-86.

Christian M, Cermak T, Doyle E L, Schmidt C, Zhang F, Hummel A, Bogdanove A J *et al.* (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* **186**:757-61. doi: 10.1534/genetics.110.120717

Datsenko K A, Pougach K, Tikhonov A, Wanner B L, Severinov K and Semenova E (2012) Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nature Commun* 3:945.

D'Halluin K, Vanderstraeten C, Van Hulle J, Rosolowska J, Brande I V D, Pennewaert A, D'Hont K *et al.* (2013) Targeted molecular trait stacking in cotton through targeted double-strand break induction. *Plant Biotechnol J* **11**:933-41. doi: 10.1111/pbi.12085

Diouf L, Pan Z, He S, Gong W F, Hua J, Yin M, Richard R, Kimbembe R, Rashid H, Kirungu J and Du X (2017) High-density linkage map construction and mapping of salt-tolerant QTLs at seedling stage in upland cotton using genotyping by sequencing (GBS). *Int J Mol Sci* **18**(12):2622.

Doudna J A and Charpentier E (2014) Genome editing: The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**:1258096. doi: 10.1126/science.1258096

Fan L, Wang L, Wang X, Zhang H, Zhu Y, Guo J, Gao W, Geng H, Chen Q and Qu Y (2018) A high-density genetic map of extra-long staple cotton (*Gossypium barbadense)* constructed using genotyping-by-sequencing based single nucleotide polymorphic markers and identification of fibre traits-related QTL in a recombinant inbred line population. *BMC Genomics* **19**:489.

Fang D D (2015) Molecular Breeding. *Cotton* 2nd Edition. David D Fang and Richard G Percy (Editors) Agronomy Monograph. **57**:255-88.

Fang D D and Yu J Z (2012) Addition of 455 microsatellite marker loci to the high-density *Gossypium hirsutum* TM-1 x *G. barbadense* 3-79 genetic map. *J Cotton Sci* **16**:229-48.

Fang L, Gong H, Hu Y, Liu C, Zhou B, Huang T, Wang Y *et al.* (2017a) Genomic insights into divergence and dual domestication of cultivated allotetraploid cottons. *Genome Biol* **18**:33.

Fang L, Wang Q, Hu Y, Jia Y, Chen J, Liu B, Zhang Z *et al.* (2017b) Genomic analyses in cotton identify signatures of selection and loci associated with fibre quality and yield traits. *Nat Genet* **49**:1089-98.

Feng X, Keim D, Wanjugi H, Coulibaly I, Fu Y, Schwarz J, Huesgen S *et al.* (2015) Development of molecular markers for genetic male sterility in *Gossypium hirsutum*. *Mol Breeding* **35**:141.

Fire A, Xu S, Montgomery M K, Kostas S A, Driver S E and Mello C C (1998) Potent and specific genetic interference by double stranded RNA in *Caenorhabditis elegans*. *Nature* **19**:806-11.

Gallagher J P, Grover C E, Rex K, Moran M and Wendel J F (2017) A new species of cotton from Wake Atoll, *Gossypium stephensii* (malvaceae). *Systematic Bot* **42**:115-23.

Gao W, Long L, Tian X Q, Xu F C, Liu J, Singh P K, Botella J R *et al.* (2017) Genome editing in cotton with the CRISPR/Cas9 system. *Front Plant Sci* **8**:1364.

Gomez G M, Moretzon M C, Da Silva E H, Cleber F, Silvba J G P, Lacape J M, Vianna B P A *et al.* (2016) Mapping of QTLs identified in a novel source of resistance to the Root-Knot nematode (*Meloidogyne incognita* race 3) in cotton (*Gossypium barbadense* L.). Goiânia: ICGI, Résumé, 2 p. World Cotton Research Conference. 6, Goiânia, Brésil, 2-6 May 2016 (Abstr).

Gore M A, Fang D D, Poland J A, Zhang J, Percy R G, Cantrell R G, Thyssen G *et al.* (2014) Linkage map construction and quantitative trait locus analysis of agronomic and fibre quality traits in cotton. *Plant Gen* **7**(1). doi: 10.3835/plantgenome2013.07.0023

Grover C E, Grupp K K, Wanzek R J and Wendel J F (2012) Assessing the monophyly of polyploid *Gossypium* species. *Plant Syst Evol* **298**:1177-83. doi: 10.1007/s00606-012-0615-7

Hagenbucher S, Eisenring M, Meissle M, Rathore K S and Romeis J (2019) Constitutive and induced insect resistance in RNAi-mediated ultra-low gossypol cottonseed cotton. *BMC Plant Biol* **19**:322. doi: 10.1186/s12870-019-1921-9

Handi S S, Katageri I S, Adiger S, Jadhav M P, Lekkala S P and Lachagari V B R (2017) Association mapping for seed cotton yield, yield components and fibre quality traits in upland cotton (*Gossypium hirsutum* L.) genotypes. *Plant Breeding* doi: 10.1111/pbr.12536

Hendrix B and Stewart J M (2005) Estimation of the nuclear DNA content of *Gossypium* species. *Ann Bot (Lond.)* **95**:789-97. doi: 10.1093/aob/mci078

Hinze L L, Hulse-Kemp A M, Wilson I W, Zhu Q H, Llewellyn D J, Taylor J M, Spriggs A *et al.* (2017) Diversity analysis of cotton (*Gossypium hirsutum* L.) germplasm using the CottonSNP63K Array. *BMC Plant Biol* **17**:37.

Huang C, Nie X, Shen C, You C, Li W, Zhao W, Zhang X *et al.* (2017) Population structure and genetic basis of the agronomic traits of upland cotton in China revealed by a genome wide association study using high-density SNPs. *Plant Biotechnol J* **15**:1374-86.

Huang P, Xiao A, Zhou M, Zhu Z, Lin S and Zhang B (2011) Heritable gene targeting in zebrafish using customized TALENs. *Nat Biotechnol* **29**:699-700. doi: 10.1038/nbt.1939

Hulse-Kemp A M, Lemm J, Plieske J, Ashrafi H, Buyyarapu R, Fang D D, Frelichowski J *et al.* (2015) Development of a 63K SNP array for cotton and high-density mapping of intra- and interspecific populations of *Gossypium spp. G3 (Bethesda)* **5**:1187-209.

Iqbal Z, Sattar M N and Shafiq M (2016) CRISPR/Cas9: a tool to circumscribe cotton leaf curl disease. *Front Plant Sci* **7**:475.

Ishino Y, Shinagawa H, Makino K, Amemura M and Nakata A (1987) Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J Bacteriol* **169**:5429-33.

Jamshed M, Jia F, Gong J, Palanga K K, Shi Y, Li J, Shang H *et al.* (2016) Identification of stable quantitative trait loci (QTLs) for fibre quality traits across multiple environments in *Gossypium hirsutum* recombinant inbred line population. *BMC Genomics* **17**:197.

Janga M R, Campbell L M and Rathore K S (2017) CRISPR/Cas9-mediated targeted mutagenesis in upland cotton (*Gossypium hirsutum* L.). *Plant Mol Biol* **94**:349-60.

Jansen R, Embden J D, Gaastra W and Schouls L M (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* **43**:1565-75.

Jia X, Pang C, Wei H, Wang H, Ma Q, Yang J, Cheng S *et al.* (2016) High-density linkage map construction and QTL analysis for earliness-related traits in *Gossypium hirsutum* L. *BMC Genomics* **17**:909.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna J A and Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**:816-21.

Khan M K, Chen H, Zhou Z, Ilyas M, Wang X, Cai X, Wang C *et al.* (2016) Genome wide SSR high density genetic map construction from an interspecific cross of *Gossypium hirsutum* × *Gossypium tomentosum*. *Front Plant Sci* **7**:436.

Khan Z, Khan S H, Mubarik M S, Sadia B and Ahmad A (2018) Use of TALEs and TALEN technology for genetic improvement of plants. *Plant Mol Biol Rep* **35**:1-19. doi 10.1007/s11105-016-0997-8

Kim Y G, Cha J and Chandrasegaran S (1996) Hybrid restriction enzymes: Zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci* **93**:1156-60.

Kirungu J, Deng Y, Cai X, Magwanga R, Zhou Z, Wang X, Wang X *et al.* (2018) Simple Sequence Repeat (SSR) genetic linkage map of d genome diploid cotton derived from an interspecific cross between *Gossypium davidsonii* and *Gossypium klotzschianum*. *Int J Mol Sci* ***19***:204.

Krapovickas A and Seijo G (2008) *Gossypium ekmanianum* (Malvaceae), algodon silvestre de la Republica Dominicana. *Bonplandia* **17**:55-63.

Kumar N V M, Katageri I S, Gowda S A, Adiger S, Yadava S K and Lachagari V B R (2019) 63K SNP chip based linkage mapping and QTL analysis for fibre quality and yield component traits in *Gossypium barbadense* L. cotton. *Euphytica* **215**:6.

Kushanov F, Buriev Z, Shermatov S, Turaev O, Norov T, Pepper A, Saha S *et al.* (2017) QTL mapping for flowering-time and photoperiod insensitivity of cotton *Gossypium darwinii* Watt. *PLoS One* **12**:e0186240.

Li C, Dong Y, Zhao T, Li L, Li C, Yu E, Mei L *et al.* (2016) Genome-wide SNP linkage mapping and QTL analysis for fibre quality and yield traits in the upland cotton recombinant inbred lines population. *Front Plant Sci* 7:1356.

Li C, Fu Y, Sun R, Wang Y and Wang Q (2018b) Single-locus and multi-locus genome-wide association studies in the genetic dissection of fibre quality traits in upland cotton (*Gossypium hirsutum* L.) *Front Plant Sci* **9**:1083.

Li C, Unver T and Zhang B (2017) A high-efficiency CRISPR/Cas9 system for targeted mutagenesis in Cotton (*Gossypium hirsutum* L.) *Sci Rep* **7**:43902. doi: 10.1038/srep43902

Li C, Wang Y, Ai N, Li Y and Song J (2018a) A genome-wide association study of early-maturation traits in upland cotton based on the CottonSNP80K array. *J Integr Plant Biol* **10**: 970-85.

Li F, Fan G, Lu C, Xiao G, Zou C, Kohel R J, Ma Z *et al.* (2015) Genome sequence of cultivated upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nat Biotechnol* **33**:524-30.

Li F, Fan G, Wang K, Sun F, Yuan Y, Song G, Li Q *et al.* (2014) Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nat Genet* **46**:567-72.

Li J F, Norville J E, Aach J, McCormack M, Zhang D, Bush J, Church G M *et al.* (2013) Multiplex and homologous recombination-mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9. *Nat Biotechnol* **31**:688-91.

Li L, Zhao S, Su J, Fan S, Pang C, Wei H, Wang H *et al.* (2017) High-density genetic linkage map construction by F_2 populations and QTL analysis of early-maturity traits in upland cotton (*Gossypium hirsutum* L.). *PLoS One* **12**: e0182918.

Li P, Kirungu J, Lu H, Magwanga R, Lu P, Cai X, Zhou Z *et al.* (2018) 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* **13**: e0207271.

Li S, Chen Z, Zhao N, Wang Y, Nie H and Hua J (2018) The comparison of four mitochondrial genomes reveals cytoplasmic male sterility candidate genes in cotton. *BMC Genomics* **19**:775.

Li T, Ma X, Li N, Zhou L, Liu Z, Han H, Gui Y *et al.* (2017) Genome-wide association study discovered candidate genes of Verticillium wilt resistance in upland cotton (*Gossypium hirsutum* L.). *Plant Biotech J* **15**(12):1520-32.

Li X, Jin X, Wang H, Zhang X and Lin Z (2016) Structure, evolution, and comparative genomics of tetraploid cotton based on a high-density genetic linkage map. *DNA Res* **2**3(3):283-93.

Liang Q, Hu C, Hua H, Li Z and Hua J (2013) Construction of a linkage map and QTL mapping for fibre quality traits in upland cotton (*Gossypium hirsutum* L.). *Chinese Sci Bull* **58**(26):3233-43.

Liu D, Liu F, Shan X, Zhang J, Tang S, Fang X, Liu X *et al.* (2015) Construction of a high-density genetic map and lint percentage and cottonseed nutrient trait QTL identification in upland cotton (*Gossypium hirsutum* L.). *Mol Genet Genomics* **290**:1683-00.

Liu H Y, Quampah A, Chen J H, Li R, Huang Z R, He Q L, Shi C H *et al.* (2012) QTL analysis for gossypol and protein contents in upland cottonseeds with two different genetic systems across environments. *Euphytica* **188**:453-63.

Liu Q, Singh S P and Green A G (2002) High-stearic and high-oleic cottonseed oils produced by hairpin RNA-mediated post transcriptional gene silencing. *Plant Physiol* **129**:1732-43. doi: 10.1104/pp.001933

Liu R, Gong J, Xiao X, Zhang Z, Li J, Liu A, Lu Q *et al.* (2018) GWAS analysis and QTL identification of fibre quality traits and yield components in upland cotton using enriched high-density SNP markers. *Front Plant Sci* **9**:1067.

Liu R, Wang B, Guo W, Wang L and Zhang T (2011) Differential gene expression and associated QTL mapping for cotton yield based on a cDNA-AFLP transcriptome map in an immortalized F_2. *Theor Appl Genet* **123**:439-54.

Liu X, Zhao B, Zheng H J, Hu Y, Lu G, Yang C Q, Chen J D *et al.* (2015) *Gossypium barbadense* genome sequence provides insight into the evolution of extra-long staple fibre and specialized metabolites. *Sci Rep* **5**:14139. doi: 10.1038/srep14139

Long L, Guo D D, Gao W, Yang W W, Hou L P, Ma X N, Miao Y C *et al.* (2018) Optimization of CRISPR/Cas9 genome editing in cotton by improved sgRNA expression. *Plant Methods* **14**:85.

Lopez-Lavalle B, Matheson L and Brubaker B (2011) A genetic map of an Australian wild *Gossypium* C genome and assignment of homoeologies with tetraploid cultivated cotton. *Genome* **54**(9):779-94.

Lopez-Lavalle L A, Gillespie V J, Tate W A, Ellis M H, Stiller W N, Llewellyn D L and Wilson I W (2012) Molecular mapping of a new source of Fusarium wilt resistance in tetraploid cotton (*Gossypium hirsutum* L.). *Mol Breed* **30**:1181-91.

Ma J, Geng Y, Pei W, Wu M, Li X, Liu G, Li D *et al.* (2018) Genetic variation of dynamic fibre elongation and developmental quantitative trait locus mapping of fibre length in upland cotton (*Gossypium hirsutum* L.). *BMC Genomics* **19**:882.

Ma J, Pei W, Ma Q, Geng Y, Liu Y, Liu G, Liu J *et al.* (2019) QTL analysis and candidate gene identification for plant height in cotton based on an interspecific backcross inbred line population of *Gossypium hirsutum* × *Gossypium barbadense*. *Theor Appl Genet* doi: 10.1007/s00122-019-03380-7

Magwanga R O, Lu P, Kirungu J N, Diouf L, Dong Q, Hu Y, Cai X *et al.* (2018) 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 BC_2F_2 generation. *Int J Mol Sci* ***19***:1614.

Mao Y B, Tao X Y, Xue X Y, Wang L J and Chen X Y (2011) Cotton plants expressing CYP6AE14 double-stranded RNA show enhanced resistance to bollworms. *Transgenic Res* **20**:665-73. doi: 10.1007/s11248- 010-9450-1

McCallum C M, Comai L, Greene E A and Henikoff S (2000) Targeted screening for induced mutations. *Nat Biotechnol* **18**:455-57.

Zhang M, Liu J, Wei H, Pang C, Wang H, Fan S, Wang H *et al.* (2017) Genetic analysis and preliminary mapping of the photoperiod-sensitive male sterility gene ys-1 in upland cotton. *Cotton Sci* **29**(1):9-16.

Mojica F J, Díez-Villaseñor C, García-Martínez J and Soria E (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* **60**:174-82.

Moscou M J and Bgdanove A J (2009) A simple cipher governs DNA recognition by TAL effectors. *Science* **326**:1501.

Page J T, Liechty Z S, Alexander R H, Clemons K, Hulse-Kemp A M, Ashrafi H, Van Deynze A *et al.* (2016) DNA sequence evolution and rare homoeologous conversion in tetraploid cotton. *PLoS Genet* **12**: e1006012.

Palanga K K, Jamshed M, Rashid M, Gong J, Li J, Iqbal M S, Liu A *et al.* (2017) Quantitative trait locus mapping for Verticillium wilt resistance in an upland cotton recombinant inbred line using SNP-based high-density genetic map. *Front Plant Sci* **8**:382.

Paques F and Duchateau P (2007) Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. *Curr Gene Ther* **7**:49-66.

Paterson A H, Wendel J F, Gundlach H, Guo H, Jenkins J, Jin D, Llewellyn D *et al.* (2012) Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature* **492**:423-27.

Pourcel C, Salvignol G and Vergnaud G (2005) CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA and provide additional tools for evolutionary studies. *Microbiology* **151**:653-63.

Qi H, Wang N, Qiao W, Xu Q, Zhou H, Shi J, Yan G *et al.* (2017) Construction of a high-density genetic map using genotyping by sequencing (GBS) for quantitative trait loci (QTL) analysis of three plant morphological traits in upland cotton (*Gossypium hirsutum* L.). *Euphytica* **213**(4):83.

Raja D, Kumar M S, Devi P R, Loganathan S, Ramya K, Kannan N and Subramanian V (2017) Identification of molecular markers associated with genic male sterility in tetraploid cotton (*Gossypium hirsutum* L.) through bulk segregant analysis using a cotton SNP 63K array. *Czech J Genet Plant Breed* doi: 10.17221/25/2017-CJGPB

Ramesh U M, Methre R, Kumar N V M, Katageri I S, Gowda S A, Adiger S *et al.* (2019) Genome mapping and molecular mark-

ers identification for yield, yield component and fibre quality traits in tetraploid cotton. *Plant Breed* **00**:1-17.

Ravichanndran M, Deashis R, Subramanyam K and Govindaraj M (2019) Genetic linkage map construction for fibre quality traits in intraspecific upland cotton (*Gossypium hirsutum* L.). *Eurasia J Biosci* **13**:127-34.

Reinisch A J, Dong J, Brubaker C L, Stelly D M, Wendel J F and Paterson A H (1994) A detailed RFLP map of cotton, *Gossypium hirsutum* x *Gossypium barbadense*: Chromosome organization and evolution in a disomic polyploid genome. *Genetics* **138**:829-47.

Reyon D, Khayter C, Regan M R, Joung J K and Sander J D (2012a) Engineering designer transcription activator like effector nucleases (TALENs). *Curr Protoc Mol Biol* Chapter 12: Unit 12.15. doi: 10.1002/0471142727.mb1215s100

Reyon D, Tsai S Q, Khayter C, Foden J A, Sander J D and Joung J K (2012b) FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol* **30**:460-65.

Saeed M, Guo W, Ullah I, Tabbasam N, Zafar Y, Rahman M and Zhang T (2011) QTL mapping for physiology, yield and plant architecture traits in cotton (*Gossypium hirsutum* L.) grown under well-watered versus water-stress conditions. *Electron J Biotechnol* **14**:1-13.

Sekhar L, Khadi B M, Patil R S, Katageri I S and Mukri G (2016) Biochemical and molecular dissection of thermo-sensitive genetic male sterility in diploid cotton (*Gossypium arboreum* L.) *J Environ Biol* **37**:579-83.

Shaheen T, Zafar Y and Rahman M (2013) QTL mapping of some productivity and fibre traits in *Gossypium arboreum. Turk J Bot* **37**:802-10.

Shan Q, Wang Y, Li J, Zhang Y, Chen K, Liang Z, Zhang K *et al.* (2013) Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat Biotechnol* **31**:686-88.

Shang L, Liang Q, Wang Y, Wang X, Wang K, Abduweli A, Ma L, Cai S and Hua J (2015) Identification of stable QTLs controlling fibre traits properties in multi-environment using recombinant inbred lines in Upland cotton (*Gossypium hirsutum* L.). *Euphytica* **205**:877-88.

Shang L, Wang Y, Cai S, Wang X, Li Y, Abduweli A and Hua J (2016) Partial dominance, overdominance, epistasis and QTL by environment interactions contribute to heterosis in two upland cotton hybrid. *G3 (Bethesda)* **6**:499-507.

Shao Q, Zhang F, Tang S, Liu Y, Fang X, Liu D, Liu D J *et al.* (2014) Identifying QTL for fibre quality traits with three upland cotton (*Gossypium hirsutum* L.) populations. *Euphytica* **198**:43-58.

Shi Y, Zhang B, Liu A, Li W, Li J, Lu Q, Zhang Z *et al.* (2016) Quantitative trait loci analysis of Verticillium wilt resistance in interspecific backcross populations of *Gossypium hirsutum* × *Gossypium barbadense. BMC Genomics* **17**:877.

Smith J, Grizot S, Arnould S, Duclert A, Epinat J C, Chames P, Prieto J *et al.* (2006) A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. *Nucleic Acids Res* **34**:e149.

Song C, Li W, Pei X, Liu Y, Ren Z, He K, Zhang F *et al.* (2019) Dissection of the genetic variation and candidate genes of lint percentage by a genome-wide association study in upland cotton. *Theor Appl Genet* **132**:1991-2002.

Sun F D, Zhang J H, Wang S F, Gong W K, Shi Y Z, Liu A Y, Li J W *et al.* (2012) QTL mapping for fibre quality traits across multiple generations and environments in upland cotton. *Mol Breed* **30**:569-82.

Sun Z, Wang X, Liu Z, Gu Q, Zhang Y, Li Z, Ke, H *et al.* (2017) Genome-wide association study discovered genetic variation and candidate genes of fibre quality traits in *Gossypium hirsutum* L. *Plant Biotechnol J* **15**:982-96.

Sunilkumar G, Campbell L M, Puckhaber L, Stipanovic R D and Rathore K S (2006) Engineering cottonseed for use in human nutrition by tissue specific reduction of toxic gossypol. *Proc Natl Acad Sci USA* **103**:18054-59. doi: 10.1073/pnas.0605389103

Tan Z, Fang X, Tang S, Zhang J, Liu D, Teng Z, Li L *et al.* (2015) Genetic map and QTL controlling fibre quality traits in upland cotton (*Gossypium hirsutum* L.). *Euphytica* **203**:615-28.

Tan Z, Zhang Z, Sun X, Li Q, Sun Y, Yang P, Wang W *et al.* (2018) Genetic map construction and fibre quality QTL mapping using the cotton SNP 80K array in upland cotton. *Front Plant Sci* **9**:225.

Tang S, Teng Z, Zhai T, Fang X, Liu F, Liu D, Zhang J *et al.* (2015) Construction of genetic map and QTL analysis of fibre quality traits for Upland cotton (*Gossypium hirsutum* L.). *Euphytica* **201**(2):195-213.

Trapero C, Wilson I W, Stiller W N and Wilson L J (2016) Enhancing integrated pest management in GM cotton systems using host plant resistance. *Front Plant Sci* **7**:500.

Ulloa M, Hulse-Kemp A M, De Santiago L M, Stelly D M and Burke J J (2017) Insights Into Upland Cotton (*Gossypium hirsutum* L.) Genetic recombination based on 3 high-density single-nucleotide polymorphism and a consensus map developed independently with common parents. *Genomics Insights* **10**:1-15.

Wang B, Liu L, Zhang D, Zhuang Z, Guo H, Qiao X, Wei L *et al.* (2016) A genetic map between *Gossypium hirsutum* and the Brazilian endemic *G. mustelinum* and its application to QTL mapping. *G3 (Bethesda)* **6**(6):1673-85.

Wang H, Huang C, Guo H, Li X, Zhao W, Dai B, Yan Z *et al.* (2015a) QTL mapping for fibre and yield traits in upland cotton under multiple environments. *PLoS One* **10**: e0130742.

Wang H, Jin X, Zhang B, Shen C and Lin Z (2015b) Enrichment of an intraspecific genetic map of upland cotton by developing markers using parental RAD sequencing. *DNA Res* **22**:147-60.

Wang K, Wang Z, Li F, Ye W, Wang J, Song G, Yue Z *et al.* (2012) The draft genome of a diploid cotton *Gossypium raimondii*. *Nat Genet* **44**:1098-103.

Wang K, Wendel J F and Hua J (2018) Designations for individual genomes and chromosomes in *Gossypium*. *J Cotton Res* **1**:3. doi: 10.1186/s42397-018-0002-1

Wang M, Li C and Wang Q (2014) Quantitative trait loci mapping and genetic dissection for lint percentage in upland cotton (*Gossypium hirsutum*). *J Genet* **93**(2):371-78.

Wang M, Tu L, Lin M, Lin Z, Wang P, Yang Q, Ye Z *et al.* (2017) Asymmetric subgenome selection and cis-regulatory divergence during cotton domestication. *Nat Genet* **49:**579-87.

Wang P Z, Bian L and Cao C (2016) Mapping of nuclear male-sterile gene ms14 using SSR markers in cotton. *Mol Plant Breed* **7**(35):1-5.

Wang P, Zhang J, Sun L, Ma Y, Xu J, Liang S, Deng J *et al.* (2018) High efficient multi-sites genome editing in allotetraploid cotton (*Gossypium hirsutum*) using CRISPR/Cas9system. *Plant Biotechnol J* **16**:137-50.

Wang X, Yu Y, Sang J, Wu Q, Zhang X and Lin Z (2013) Intraspecific linkage map construction and QTL mapping of yield and fibre quality of *Gossypium babardense*. *Aust J Crop Sci* **7**(9):1252-61.

Wang Y L, Meng Z G, Liang C Z, Meng Z H, Wang Y, Sun G Q, Zhu T *et al.* (2017) Increased lateral root formation by CRISPR/Cas9-mediated editing of arginase genes in cotton. *Sci China Life Sci* **60**:524-27.

Wang Y, Ning Z, Hu Y, Chen J, Zhao R, Chen H, Ai N *et al.* (2015) Molecular mapping of restriction-site associated DNA markers in allotetraploid upland cotton. *PLoS One* **10**: e0124781.

Weber E, Gruetzner R, Werner S, Engler C, Marillonnet S (2011) Assembly of designer TAL effectors by golden gate cloning. *PLoS ONE* **6**: e19722. doi: 10.1371/journal.pone.0019722

Wendel J F and Cronn R (2003) Polyploidy and the evolutionary history of cotton. *Adv Agron* **78**:139-86.

Xu Z, Zhang C, Ge X, Wang N, Zhou K, Yang X, Wu Z *et al.* (2015) Construction of a high-density linkage map and mapping quantitative trait loci for somatic embryogenesis using leaf petioles as explants in upland cotton (*Gossypium hirsutum* L.). *Plant Cell Rep* **34**:1177-87.

Yu J Z, Kohel R J, Fang D D, Cho J, Deynze AV, Ulloa M, Hoffman S M *et al.* (2012) A high-density simple sequence repeat and single nucleotide polymorphism genetic map of the tetraploid cotton genome. *G3 (Bethesda)* **2**(1):43-58.

Yu J, Zhang K, Li S, Yu Y, Zhai H, Wu M, Li X *et al.* (2013) Mapping quantitative trait loci for lint yield and fibre quality across environments in a *Gossypium hirsutum* x *Gossypium barbadense* backcross inbred line population. *Theor Appl Genet* **126**:275-87.

Yu Y, Yuan D J, Liang S G, Li X M, Wang X Q, Lin Z X and Zhang X L (2011) Genome structure of cotton revealed by a genome wide SSR genetic map constructed from a BC1 population between *Gossypium hirsutum* and *G. barbadense*. *BMC Genomics* **12**:15.

Zhang C, Li L, Liu Q, Gu L, Huang J, Wei H, Wang H *et al.* (2019) Identification of loci and candidate genes responsible for fibre length in upland cotton (*Gossypium hirsutum* L.) via association mapping and linkage analyses. *Front Plant Sci* **10**:53.

Zhang J, Yu J, Pei W, Li X, Said J, Song M and Sanogo S (2015) Genetic analysis of Verticillium wilt resistance in a backcross inbred line population and a meta-analysis of quantitative trait loci for disease resistance in cotton. *BMC Genomics* **16**:577.

Zhang K, Zhang J, Ma J, Tang S Y, Liu D J, Teng Z H, Liu D X *et al.* (2012) Genetic mapping and quantitative trait locus analysis of fibre quality traits using a three-parent composite population in upland cotton (*Gossypium hirsutum* L.). *Mol Breed* **29**:335-48.

Zhang S, Feng L, Xing L, Yang B, Gao X, Zhu X, Zhang T *et al.* (2016) New QTLs for lint percentage and boll weight mined in introgression lines from two feral landraces into *Gossypium hirsutum* acc TM-1. *Plant Breed* **135**: 90-101.

Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, Zhang J *et al.* (2015) Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fibre improvement. *Nat Biotechnol* **33**:531-37.

Zhang Z, Ge Q, Liu A, Li J, Gong J, Shang H, Shi Y *et al.* (2017) Construction of a high-density genetic map and its application to QTL identification for fibre strength in upland cotton. *Crop Sci* **57**:774-88.

Zhao C, Zhao G, Geng Z, Wang Z, Wang K, Liu S, Zhang H *et al.* (2018) Physical mapping and candidate gene prediction of fertility restorer gene of cytoplasmic male sterility in cotton. *BMC Genomics* **19**:6.

Zhao J, Liu J, Xu J, Zhao L, Wu Q and Xiao S (2018) Quantitative trait locus mapping and candidate gene analysis for *Verticillium* wilt resistance using *Gossypium barbadense* chromosomal segment introgressed line. *Front Plant Sci* **9**:682.

Zhao L, Lv Y, Cai C, Tong X, Chen X, Zhang W, Du H *et al.* (2012) Towards allotetraploid cotton genome assembly: Integration of a high-density molecular genetic linkage map with DNA sequence information. *BMC Genomics* **13**:539.

Zhao Y, Wang H, Chen W, Zhao P, Gong H, Sang X and Cui Y (2017) Regional association analysis-based fine mapping of three clustered QTL for verticillium wilt resistance in cotton (*G. hirsutum.* L). *BMC Genomics* **18**:661.

Zhiyuan N, Chen H, Mei H and Zhang T (2014) Molecular tagging of QTLs for fibre quality and yield in the upland cotton cultivar Acala-Prema. *Euphytica* **195**:143-56.

Zhu Q H, Yuan Y, Stiller W, Jia Y, Wang P, Pan Z, Du X *et al.* (2018) Genetic dissection of the fuzzless seed trait in *Gossypium Barbadense. J Exp Bot* **69**(5):997-1009.

Zhu Q H, Zhang J, Liu D, Stiller W, Liu D, Zhang Z, Llewellyn D *et al.* (2016) Integrated mapping and characterization of the gene underlying the okra leaf trait in *Gossypium hirsutum* L. *J Exp Bot* **67**:763-74.

Zhu S, Yu X, Li Y, Sun Y, Zhu Q and Sun J (2018) Highly efficient targeted gene editing in upland cotton using the CRISPR/Cas9 system. *Int J Mol Sci* **19**:3000. doi: 10.3390/ijms19103000.

The ICAC Recorder

Mini-Review

Innovative Approaches to Breeding Cotton

Shreekant S. Patil, Former Director of Research . University of Agricultural Sciences (UAS), Dharwad - 580005, India, E-mail: sspadvance@gmail.com

Indian cotton yields have been stagnant at 437 to 566 Kg/hectares (ha) from 2004 to 2018. The 15-years (2004 to 2018) average yield (511 Kg/ha) of Indian cotton is 44% less than the 15-years average yield (905 Kg/ha) of rest of the world (Kranthi, ICAC, personal communication). Cotton in north India is grown under almost completely irrigated conditions on about 1.5 million ha. Even the 15-year average yield of 567 Kg/ha in irrigated north India is 37% less than the 15-year average yield of rest of the world, where only 45% of cotton is grown under irrigation (Kranthi, ICAC, personal communication). Indian cotton yield ranks below at least 32 countries. The low rank is despite hybrid technology, double-gene-Bt, high fertiliser use, 35% irrigated area, access to pesticides and other technologies — advantages that many African countries do not have, yet they are ahead of India in yields. This paper explores innovative approaches to breeding cotton that can catalyse high yields in India with lowered production costs.

Cotton is a unique commercial crop, which constantly encounters an array of intimidating problems that bring down the remunerative value of the crop. There are problems of biotic stresses like bollworms, sucking pests and viruses, which constantly evolve and adapt to break down the barriers of host-plant resistance. Agriculture, especially cotton cultivation in the Indian subcontinent, still continues to be labour intensive and least mechanised. The current bushy type of plant architecture in India is not well suited for mechanical planting and harvesting. Further, the planting density in India is low and the crop season extends over 5-8 months, which requires more labour for crop production and multiple pickings. The increasing labour costs in cotton cultivation result in declining returns. It has now become necessary to explore options to alter the plant type to make it as compact as possible so as to increase the planting density, condense the growing season to 5 months and make it amenable for mechanised planting and harvesting. Compact cottons have the primary advantage of planting in high density thereby contributing to high yields coupled with mechanised cotton cultivation, reduced intensity of pest damage due to a shorter season and facilitating a double-cropping system for ecological sustainability and enhanced remuneration in cotton-based cropping systems.

Interestingly, even the Indian seed industry has been contemplating a change in its mindset to develop short-season Bt hybrids with compact architecture due to their ability to escape the late season pink bollworm infestation and their suitability for machine sowing and picking. However, there is a practical limit up to which compactness can be introduced in hybrids to reduce spacing and increase density, as it leads to increase in hybrid seed cost. In contrast, it is possible to introduce a high degree of compactness in varietal background to easily increase the density to more than 100,000 plants per hectare. It is thus possible to reap the benefit of varietal compactness to a much greater extent than compactness in hybrids. It is necessary for public sector institutions to focus research on developing gradations of compact types and determine the ideotype of cotton for different growing conditions in India. There is a need for research work to determine the optimum height of the productive compact cotton varieties grown for manual picking at present and for machine picking to be followed in future under different situations. Presently, cotton is manually picked in the Indian subcontinent and a plant height of 6-8 feet is common. For machine picking, the plant dimensions are restricted to 3 feet high and 2 feet wide so that the plants do not break when they pass through the spindle assembly window. Thus, the future plant types being developed for high density planting should be suitable for machine sowing and there should be a scientific debate among breeders on the ideal height for manual picking and whether tall compact plant types can out-yield dwarf compact types under different crop production systems in India. In any case, the plant type should be accessible for weeding, spraying with tractor and such other forms of mechanised operations. Importantly, the cost of varietal seed in high density is much lower compared to hybrids apart from making farmers self-sufficient for their seed needs.

There are several challenges with breeding for compact plant types. While it has taken several years for countries such as Australia, Brazil, China, Mexico, Turkey, Uzbekistan and USA to develop varieties with compact architecture and premium quality fibres, efforts in India are in their infancy. Modernisation of textile industrial operations constantly raises the benchmark of fibre quality requirements. When breeders in India opt for developing compact types with earliness, they still wish the plant to retain a larger boll size for convenience of manual picking, coupled with high fibre quality and synchronous maturity. However, this involves a contradictory combination of

traits which goes against the general trend of correlations existing among them. When requirements of genetic improvement become so complex, the approaches of breeding should also match the challenge to become advanced and comprehensive.

A thorough understanding of principles of population genetics and quantitative genetics is fundamental for planning and execution of breeding approaches for creation and exploitation of genetic variability for higher genetic gains. Every breeder, even in remote research centres, addresses these complex issues in his own way and there is no system available for assessment of magnitude of useful variability released and the extent of genetic improvement achieved in the attempts. There should be a simple reporting system among breeders that incorporates the results of successes, reports of failures. assessment of reasons of failure, mistakes committed and their impact in terms of reduced genetic improvement. These studies should act as guides to other breeders in pursuing different breeding approaches in cotton and even other crops. This will help breeders learn and benefit from prior breeding attempts.

The harvesters currently used in developed countries demand the choice of compact cotton genotypes ideally suited for such machine picking. The kind of pickers chosen for use will have a bearing on the plant type suiting the situation of cotton cultivation. While research on developing plant types suited for machine pickers continues, research should also be contemplated to modify tractor drawn cotton picking machines to suit the small-medium holder conditions of India and also to develop back-pack type of pickers for the resource-poor small-holder farmers.

Genetic improvement of cotton is aimed at developing improved varieties and potential hybrids for which a sound knowledge of principles of plant breeding and genetics is fundamental to guide the various stages of genetic improvement. Lack of this leads to oversimplified procedures and defective execution of steps at different phases of varietal improvement. Due to this, though the amount of time and energy involved in the procedures for creation of variability and its exploitation remains same, realised genetic gains produced by breeders remains low. This paper dwells on explaining the genetic basis in deriving modifications in conventional breeding methods. It shows how it is necessary to extend some principal steps of breeding cross-pollinated (random mating system) crops and blending them with routine procedures of self-pollinated crops to improve the efficiency in genetic improvement of cotton for both varietal and hybrid development. Whenever possible, outcomes of our own research findings are summarised and presented in support of the concepts and in some areas, modifications in procedures/steps are suggested based only on explanation of genetic basis of the innovative approaches and modifications in steps of cotton breeding.

Genetic complementation of parents for creation of useful variability

Success of varietal improvement depends up on the magnitude of useful variability availed for a large number of traits such as yield, fibre quality and stress tolerance, followed by efficient exploitation through selection. Desirable variability for an array of these traits will be released only when the parents chosen for hybridisation are genetically diverse enough to complement for all these traits and as a result the F_1 progeny segregates for this large number of loci.

National programs for creation of pool of desirable alleles and structuring transgressive segregants blending desirable traits

Success of a national programme depends on creation of sufficient positive variability and successful exploitation of the same to derive desired potential transgressive segregants. If a single attempt of genetic improvement has to address such complex group of traits and needs of cotton plant, attempts of hybridisation for varietal improvement should systematically address complementation of the parents for the entire range of these traits. Generally, it is very difficult to get two parents that are perfectly complementary for the entire range of traits. It is necessary to choose diverse sets of parents complementing each other for a range of traits of interest and utilise them scientifically to create a wide range of variability through innovative breeder approaches and novel methods of genetic improvements. The distribution of desirable traits in a parent must determine its contribution to the pool of favourable alleles generated. Thus, the contribution of each parent to the pool of alleles could be same or different. In the development process, the expertise of entomologists, physiologists and agronomists can be crucial in determining the score card of each genotype in terms of desirable expression for traits of economic importance like yield, its trait components, fibre quality, stress tolerance, biomass features, harvest index and relative growth rate.

When such an integrated attempt is made, it is possible that variability is created for a very large number of loci with huge variability for isolation of potential transgressive segregants blending a big list of desired traits. The huge magnitude of variability released by the large sets of parents cannot be managed by individual breeders. Therefore, concerted efforts should be made regularly through national programs to create enormous positive variability by supporting a large group of breeders identified in a region who would screen the large populations and distribute them for practicing selection and exploiting the same for genetic improvement.

Choice of a set of parents for hybridisation can be effectively done by assessing the score card of genotypes for an

array of requisite traits to decide upon the group of parents that can complement each other for the entire range of traits. It is difficult to expect just two parents to perfectly complement each other for the entire range of yield components either influencing biomass or translocation of photosynthates to sink, (as shown in Fig. 1) biotic, abiotic stress related traits and an array of other desired traits.

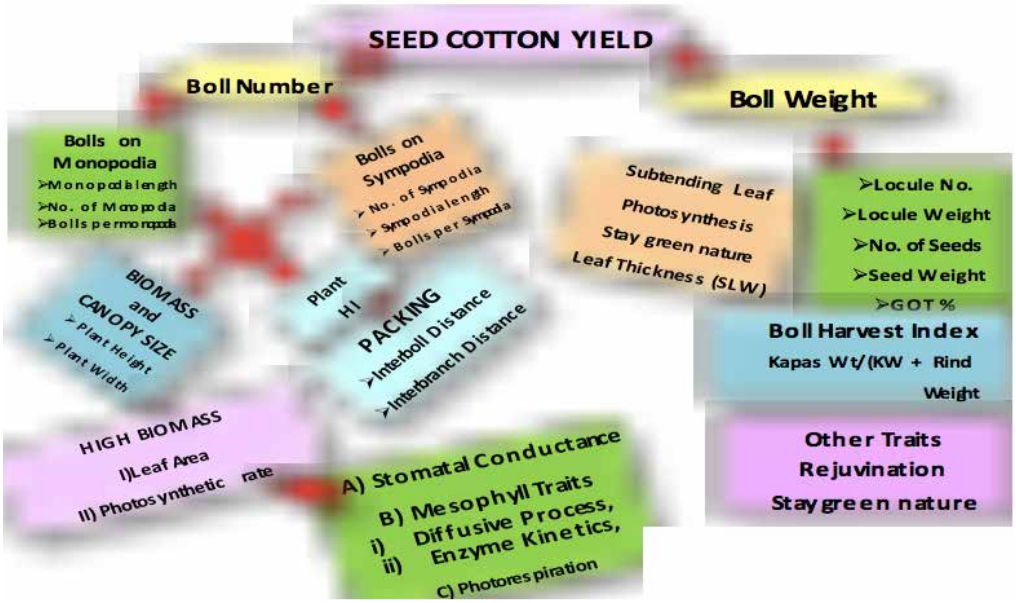


Figure 1. Path of productivity in cotton

Efficiency of releasing useful variability improves when multiple parents accounting for the desirable expression for an array of characters are used as they account for better complementation of an array of important traits as compared with the use of just two parents.

We studied variability in different combinations of four carefully chosen diverse parents representing genetically diverse (heterotic) groups to compare with that based on two parents. For this purpose, four parents were used from proven plant type diversity groups such as robust, stay green types on one side and compact, high relative growth rate (RGR) types on other side. It was observed that the proportion of transgressive segregants was much higher when such four-parent based segregating populations were developed for practicing selection in segregating generations (Edke, 2016). These results clearly indicate that for effective varietal improvement (genetic gain) in cotton, the existence of diversity of parents and complementation for the entire range of traits is very essential.

Effective modifications in approaches of handling segregating populations

After creation of useful variability, the next dimension of factors operates at exploitation of the variability created through hybridisation. For this, it is essential to recollect some basic concepts related to the consequences of selfing.

Consequences of selfing and constitution of selfed generations

Simple Mendelian expectations of segregation at a locus form the basis for determining the constitution of plants in the segregating populations derived through selfing. A heterozygous F_1 segregates to give 2^n gametic types and 3^n genotypic classes. As the number of loci influencing the ultimate dependent character like yield goes on increasing, values of total number of genotypes produced in F_2 generation reaches astounding figures. Breeders should be able to distinguish between a minimum population size and an optimum population size to be raised in the F_2 generation. It is interesting to know how this minimum expected population size changes over different segregating generations. When an F_1 segregates for a meagre 21 loci over 2 million gametic types are produced by F_1, and when these gametes unite, 10 billion genotypes are produced and this demands a minimum population of 4 trillion plants raised in F_2 generation and this is far less than the optimum population size required to be raised for these many loci. Though Allard (1960) addresses the issue of minimum population size, no mention is made about the optimum population size required to be grown. It is difficult to achieve optimum plant population size even when a F_2 population is segregating even for 21 loci? If a realistic number of loci (hundreds) governing inheritance of a dependent character like yield are considered, the genotypic classes and minimum population size reach astronomical figures and the population size normally used for handling F_2 segregating generation becomes abysmally insufficient to capture the variability. If the entire range of variability is not captured, it is difficult to expect the most potential genotypes to be seen in small population raised by a breeder. The cartoon (Fig. 2) highlights how small population size becomes just incompetent to capture the full range of variability released in F_2 generation obtained from a cross of truly diverse parents (segregating for hundreds of loci).

Figure 2. Cartoon depicting contradictions about optimum plant population and actual plant population on which a breeder exercise selectionwith an aim to develop most ideal variety.

National programmes consisting of team of breeders and supporting scientists

Many of the problems of cotton are too complex to be tackled by individual breeder and therefore there is a dire need for national level programs aimed at concerted efforts of team of breeders across different centres and supported by scientists of other related disciplines like entomology, pathology, physiology, agronomy, etc., so that the seeds of the broad-based populations developed in this manner are distributed to individual breeders for selection and stabilisation.

Such programs can generate useful data. One of the objectives of research could be to determine what happens when the F_2 population is shared and distributed to number of breeders at various locations so that its true potentiality is realised somewhere by some breeder compared to the results obtained by a single breeder — who would have created a huge useful variability but ultimately would have struggled to handle the transgressive segregants. There is a need for developing a yardstick for determination of the acceptable optimum population size based on the visible diversity among the parents for different yield related traits. It is essential that breeders exercise modifications in procedures of handling segregating generations along with a team of subject experts like entomologists, agronomists, pathologists for help in determining the target genotype and to ensure higher success rate.

Thus far, there are hardly any studies conducted to determine the minimum F_2 population of cotton that should be raised and to evaluate the possible impact of a small sample size (a few hundred plants) on the reduction in observed variability and loss of transgressive segregants. We compared the variability and frequency of transgressive segregants out of 2000 plants by taking a random sample of 200 plants each time out of this reasonable F_2 population for recording observations on productivity traits. Our data show how the most potential plants are often missed when a small population sample is examined. If the cross is segregating for realistic number of loci — hundreds, for example — it is impossible to capture the essence of the total variability released by the cross.

As an outcome of this study, it is suggested that whenever diverse parents are used in hybridisation and large amount of variability is released, individual breeders will not be able to handle such a large F2 population for capturing the most potential transgressive segregants. Instead, if such plans are executed through national programmes and the segregating population can be shared to groups of breeders working at different locations. This will potentially increase the chance of identifying the most potential segregants which can later be shared with the breeders of the region. In this manner, such segregating populations of planned diverse crosses can be exploited better by a team of breeders.

Innovative steps to recover the most potential transgressive segregants

The genetic explanation amply reinforces the fact that just as the blue whale is missing in the bucket of the breeder, the most potential transgressive segregants are missed from the small F_2 population. Now the condition of the cotton breeder is comparable to the fate of the passenger who reaches the railway station with heavy luggage just to discover that the train has already left the station. The challenge of catching the missed train could be compared to the challenge of creating the most potential transgressive segregants.

If the F_2 population is sufficiently large — thousands — the visible most productive segregants carry many of the desirable alleles distributed among them. This is comparable to a situation when an ardent fan of a famous celebrity is ridiculed by tearing the photo of the celebrity star into pieces (Fig. 3) and the fan out of dedication collects these pieces to recreate the lost picture to the extent possible. Only when the productive segregants are recombined, it is possible to recreate most potential segregants which were

Figure 3. Recombination of characters in F_2 populations and their recreation by intermating promising F_2 plants.

missed in segregating generation because of small population size raised.

In a study conducted at Dharwad on a segregating F_2 generation of *G. hirsutum*_varietal cross, the top 20 plants were advanced to F_3 in the off-season and were recombined through simulated random mating. The recombined population resulting from recombining the genetic backgrounds of these progenies was evaluated along with the normal selected plants progenies of the cross and normal (unselected) population. The mean, variance and frequency of transgressive segregants observed in this churned (recombined) population was found to be improved over the population derived by just selecting for high yield and the random population. This highlights the need for a mandatory step of recombination among the available potential segregants in early segregating generations to generate the missing most potential transgressive segregants.

Modifications in the procedures of handling segregating generations of cotton crosses

Depending on the extent of genetic divergence encountered by the breeder, it is possible to plan modifications in procedures for creation of variability and to exploit the same through selection by appropriate handling of segregating populations. Some of the modified procedures are provided below along with their genetic basis.

Genetic basis of some modifications

By F_6 generation, the minimal population size reduces from 4^n to 2^n per locus which gives a huge advantage for a large scale of segregation. Therefore, a method such as single seed descent (SSD) gets a blanket benefit over pedigree/bulk methods of breeding. There is a need to compare the efficiency of SSD over other methods in terms of reduction in minimum population size. It is suggested to delay the selection up to F_3 generation and F_2 population can be forwarded to F_3 generation exploring the option of off-season, wherever available. Postponing the artificial selection just by one generation (to F_3) provides a big advantage, considering the large number of genes for which the population is segregating as, the minimum population to be evaluated gets reduced to 3^n per locus.

Combination of SSD and pedigree methods of breeding

The main advantage of SSD is the reduction in population size to be evaluated for selection. When segregating generations are handled, minimum population size reduces from 4^n in F_2 to 2^n by about F_6 and later. This opens up possibilities of series of modifications of handling selfed segregating generations. Even if selection is postponed to F_3 by raising F_2 in off season through close spacing or broadcasting method of sowing and advancing F_2 to F_3 the minimum population size reduces approximately to 3^n in F_3. Intermating among the selected promising F_3 plants may be attempted to recreate the potential genetic variability which could have been achieved by evaluating large F_2 population. This greatly increases the chance of obtaining the most potential transgressive segregants. By following this deviation, it is possible to save one season and even gain a big advantage of reducing minimum population size.

Exploring the nature of gene action for higher genetic gains

It is possible to exploit both additive and complete dominance gene action in often-self-pollinated crops like cotton. To exploit dominance gene action, it is necessary to have lesser heterozygosity in plants to be selected. Hence, it is desirable to avoid selection in early generation like F_2 to F_4. By doing so, the deceptive level of genotypes in general reduces. This becomes an additional advantage apart from the advantage of reduced minimum population size.

Initial generations through SSD

The initial two segregating generations (F_2 and F_3) resulting from selfing show higher rise in inbreeding coefficient (to 0.75) thereby indicating that it is beneficial to follow SSD in these two early generations by raising very large population since the aim is to get just one seed from each segregant in these F_2 and F_3 generations. Artificial selection can very well be initiated in F_4 generation and all the benefits of reducing minimum size and associated benefits of starting artificial selection in F_4 can be exploited.

Alternate generations through SSD

Another alternate procedure is to grow even numbered segregating generations in off season and odd numbered segregating generations in the regular season. These procedures also ensure quickest advancing to later generations and exploiting the benefit of both pedigree and bulk method of breeding. Here, it is necessary that during off-season while advancing the population, identity is maintained for each plant by proper labelling. The F_2 population, in particular, needs to be raised during the off-season as the inbreeding coefficient increases rapidly.

Representing two generations in an evaluation

Using seeds of plants in regular season and off-season of each year, one can explore evaluation of two generations in a single row (season). Here the seeds of earlier generation (say F_3) may be used for evaluation while, at the end of each row, a small extended row (corresponding progeny of F_3) may be planted (Fig. 4). Based on performance, superior F_3 row lines can be selected but corresponding F_4 row may be harvested for advancing it to F_5 for evaluation. One can also explore to evaluate both F_4 and F_5 rows of selected lines in next year as this will help in confirming the performance of the selected lines with two generations grown in

a single season per year. The suggested modifications have general value in all self-pollinated crops and cotton breeders should make use of these innovative modifications by exercising care in handling segregating generations.

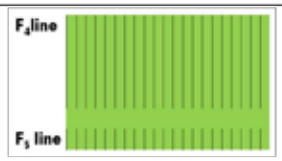
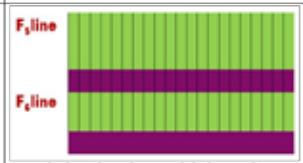
Year	Regular season	Off-season
2015	Evaluation of F_1 generation	Broadcast F_2 seeds to raise large population
2016	Evaluation of F_3 generation	F_4 generation through SSD
2017		F_6 advanced from selected F_5 lines identified based on the performance of F_4 lines
2018		With every passing year, minimum population to be grown for selection of most productive segregants decreases significantly

Figure. 4. Evaluation of two generations in the same season for improving selection efficiency

Genetic basis of backcrossing

We uncovered many interesting facts upon our detailed studies on the constitution of segregating populations as compared to constitution of backcross derived populations. The proportion of homozygous genotypes is equal to proportion of plants resembling recurrent parent. Hence, the proportion of plants which are exactly like recurrent parent is still very less in even BC_5 generation. It is the proportion of plants nearly like recurrent parent that becomes more than 90% of the BC_1 population.

By deciphering similarities and differences in consequences of selfing and backcrossing, it is possible to show that the size of backcross populations normally involved in backcrossing is very small and this can adversely affect the recovery of constitution of backcross derived generations. The minimum population size is 2^n in BC_1 generation. It works out to be 2 million plants if the genetic background is definable by just 21 genes and considering a large genetic background (n in hundreds), this becomes a very large population size. As against this requirement, the commonly used size of a BC_1 population for backcrossing is much smaller. Due to sampling effect, it can lead to drift in allelic frequencies which may slow down the recovery of constitution of the recurrent parent. The implications of these findings are that the number of plants carrying the trait under transfer from donor parent have to be large and indeed this becomes the BC population which must account for the entire range of variability released in the population. During regular backcross breeding programs of cotton meant either for transferring Bt genes or for other traits, whenever very few plants are used in different backcross generations, the introgressed or improved version show differences from the original version of recurrent parent used in backcrossing. It should be noted that the effect mentioned here is specifically for the background genes (loci) of the genotypic constitution which are not linked to the gene under transfer. The problem of delay caused in recovery of constitution of recurrent parent adds another dimension to the problem. As an implication of this, it is necessary for the breeder to use a large number of plants and even observe progeny row from each plant used for back crossing for its resemblance with the recurrent parent and preserve remnant backcross and selfed seeds of each plant for a possible future use.

Conceptualisation of *'target genotype'*

When hybridisation work involving the chosen parents is initiated, the breeder should have a clear perception of the constitution of the most potential transgressive segregant (aimed to become constitution of the new variety). The idea is to define such a genotype as *'target genotype'* in terms of the proportion of alleles retained from the two parents chosen for hybridisation (it can be defined in terms of multiple parents as well but for simplicity sake, the case of using only two parents in hybridisation is considered). The perception of a target genotype should be clear to the breeder and this enables the breeder to choose an appropriate breeding method and even to make suitable modifications in handling the segregating generations.

Procedures of handling segregating generations after hybridisation can be broadly grouped into a) Pedigree/ Bulk/ SSD method and b) Backcross method of breeding. This distinction between the breeding methods can be made in terms of the proportion of alleles from two parents (involved in crossing) observed among the segregants in the generations derived (after hybridisation) in these methods of breeding. Backcross method of breeding is used when the donor parent has highly undesirable genetic background except for one desirable simply inherited trait (Allard, 1960). Here, the proportion of alleles required from the two parents in the target genotype would be defined as 98:2 or 99:1. The genetic consequences of backcrossing and about when to use backcross breeding are well explained in different books on plant breeding and the procedure followed during backcross breeding precisely increases the chance of occurrence of such a target genotype.

The backcross method of breeding is considered as uniquely scientific especially because this breeding procedure facilitates and enhances the chance of occurrence of the target genotype (something like 99:1 in terms of proportions of alleles from the two parents) by modifying constitution of the base F_1 population (with 50:50 allelic proportions of the two parents) developed through hybridisation. In every backcross generation, the proportion

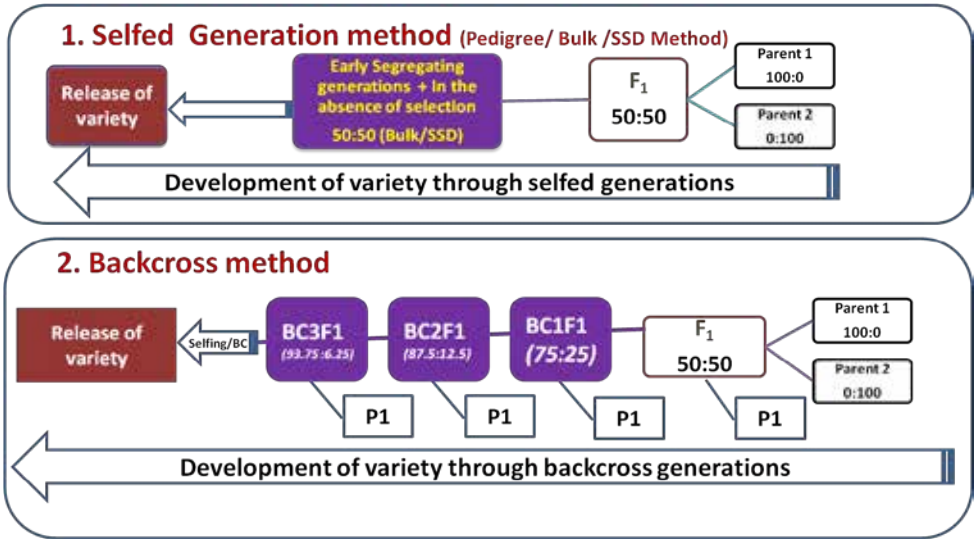


Figure 5. Handling the material after hybridisation in varietal breeding and the proportion of alleles from parents

of alleles from the recurrent parent goes on increasing and the undesirable alleles of donor parent are flushed out at high speed. If the breeder keeps track of the desirable trait of the donor parent under transfer and holds that intact in the plants chosen for back crossing, the targeted genotype is produced with great ease.

Varietal development vis-à-vis proportion of alleles from parents

With breeding methods such as Pedigree, Bulk and SSD, is it possible to expect a high frequency of target genotype? How do they compare with backcross breeding method in terms of efficiency in generating high frequency of the desired target genotype? Should these methods be rated as less scientific because the breeder is not clear about the expected target genotype set and hence thinks that it is not achieved? These are the important aspects on which a lot of information has to be generated through breeding system research, but there is a dearth of planned research in this direction. In the entire population of plants in a segregating generation, the average proportion of alleles of the two parents in F_2 and any subsequent segregating selfed generation is 50:50. It has been worked out and shown (Patil, 2014) that when genotype of each plant is examined in terms of alleles contributed by the two parents and the frequency distribution is worked out, it is observed that the proportion of plants with allelic contribution being equal to or close to 50:50 is highest (Table 1).

When this explanation is extended to cover polygenic traits, the distribution assumes the shape of normal curve. The inference derived from this distribution is that all three procedures of handling segregating generations (pedigree, bulk and SSD) will be successful when the desirable alleles for all the loci are distributed equally among the parents used for hybridisation. It also means that if the distribution of desired alleles is uneven, the pedigree method of breeding will not be successful. In contrast to this, in backcross breeding, the proportion of alleles from recurrent parent goes on increasing with every backcross generation. This is always seen as a major difference in the consequence of selfing and backcrossing. The very purpose of back crossing is to decrease the proportion of alleles of the donor parent which are associated with an undesirable genetic background. With continuous backcrossing, there is a methodical and rapid increase in the proportion of plants (in the backcross population) which resemble the recurrent parent.

The segregants most commonly observed in a breeding approach should match with the target genotype conceptualised in the beginning of varietal improvement program. Otherwise, the selected selection approach may not give the desired results. Hence, it is necessary to understand the kind of target genotype set by the particular pair of parents before embarking on the choice of breeding procedure. It is also equally important to understand the types of segregants most commonly observed in a selected breeding approach.

A segregation ratio of 1:2:1 is observed in F_2 generation with respect to a locus under consideration. When segregation at two loci is considered, a ratio of 1:4:6:4:1 is observed where the segregants with 50:50 allelic contributions from the two parents are most common (Table 1). The extreme types and those with unequal contribution of alleles from the two parents are less common. Extending it to three loci, a ratio of 1:6:15:20:15:6:1 is seen in F_2 and here again segregants with 50:50 or nearly 50:50 allelic contribution from the two parents occur most commonly in an early segregating generation. This pattern of segregation remains the same even when a large number of loci affecting a quantitative character is considered as depicted in Figure 6 and 7. In both methods of Bulk and SSD, artificial selection begins at a stage when the population is fixed or nearly fixed at the segregating loci governing inheritance of quantitative traits and at this stage the number of gametic classes and genotypic classes are same(2^n). With the help of Pascal's triangle, the segregation ratios can be worked out in F_2 generation and also in the generation representing fixation.

Table 1: Proportion of alleles of two parents among segregation ratios

a) F_2 generation

Genotypes and proportion of alleles of two parents

One gene case	AA	Aa	Aa
Proportion of alleles of parents	100:00:00	50:50:00	0,06944444
Segregation ratio	1	2	1

Genotypes and proportion of alleles of two parents

Two gene case	AABB (1)	AABb (2) AaBB (2)	AaBb (4) AAbb (1)	aaBb (2)	aabb (1)
			aaBB (1)	Aabb (2)	
Proportion of alleles of parents	100:00:00	75:25:00	50:50:00	25:75	0,06944444
Segregation ratio	1	4	6	4	1

b) F_6 generation

Genotypes and proportion of alleles of two parents

Two gene case	AABB (1)	AAbb (1)	Aabb (1)
		aaBB (1)	
Proportion of alleles of parents	100:00:00	50:50:00	0,06944444
Segregation ratio	1	2	1

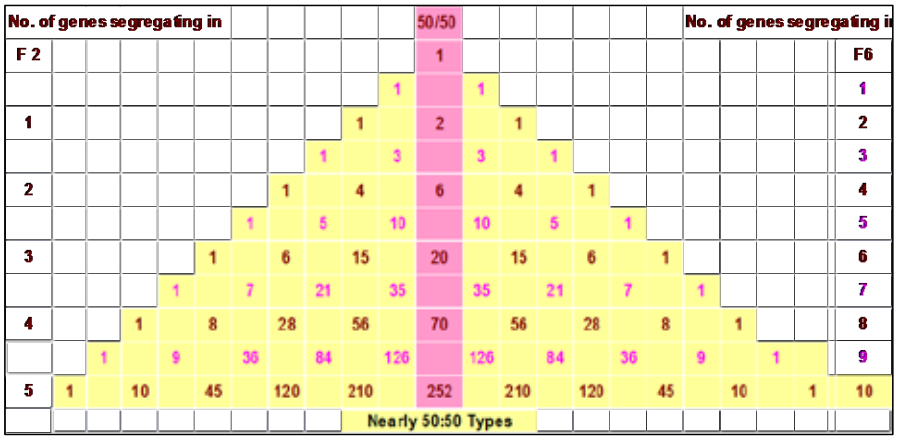


Figure 6. Proportion of alleles from the two parents among segregants in early as well as later segregating generations

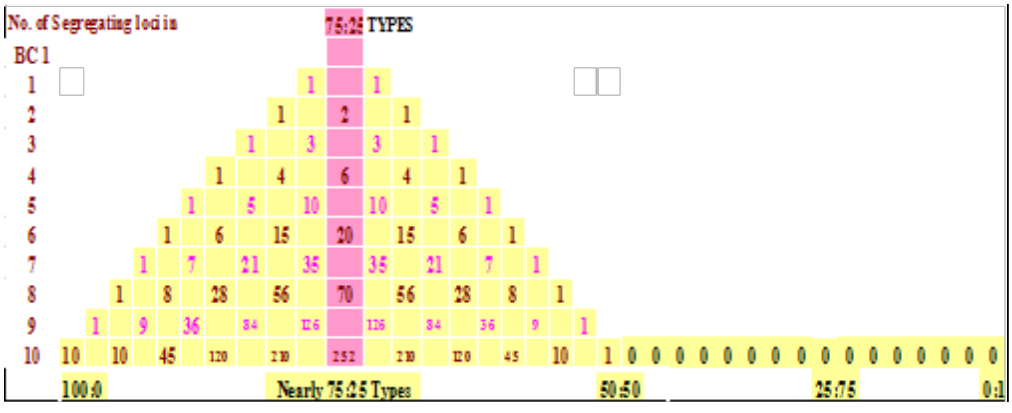


Figure 7. Proportion of alleles from the two parents among segregants in BC1 generation

In Figure 6, it is observed that the trend of prominence of 50:50 types among segregants continues even at this stage of fixation. It means that whether it is pedigree or bulk or SSD method of handling segregating populations, 50:50 types of segregants are most commonly seen in these populations subjected to artificial selection. As compared to this there is a lower frequency of 70:30 types (or symmetrically 30:70 types) and an even lesser frequency of 95:5 (or 5:95) type. It is important to remember for this reason that the target genotype achieved through back cross method of breeding (say 98:2 type) is very much available even in F_2 generation but locating such segregants in F_2 becomes a Herculean task. Therefore, it is avoided, and the tedious procedure of continuous back crossing is preferred over selection for the target 98:2 type in F_2 or later segregating generation.

Thus, the selection processes practiced in these segregating populations will be successful only if the 'target genotype' set by the pair of hybridised parents is close to 50:50. This means that if the two parents are perfectly complimenting each other on gene for gene basis or crudely speaking trait for trait (component trait) basis, the target genotype set will be 50:50 type. In simpler terms, in any segregating generation (F_2, ... F_6), the segregants which are 50:50 types (50% alleles from each parents) are the most commonly observed types. If the two parents are complementing each other perfectly by sharing the desirable alleles each at 50% of the total number of loci responsible for yield, then the target genotype in case of the parental pair chosen for hybridisation is a 50:50 type. Since the segregating population consists of mainly 50:50 types the task of the breeder is restricted to selecting the *"Best 50:50 type"*, which could be considered as the target recom-

binant type. This incidentally it is a positive transgressive segregant which blends the desirable favourable alleles equally distributed between the two parents. Apart from these extreme positive transgressive segregants among this wide array of 50:50 types of segregants, one can also expect extreme negative transgressive segregants, which may perfectly blend only the undesirable alleles distributed equally from the two parents. Since the population has high frequency of 50:50 types, selecting the best 50:50 among them becomes relatively easy. Based on this theory, it is expected that these three methods of handling segregating populations after hybridisation will be successful only when the target genotype is a 50:50 type. It should also be remembered that following these methods will be ineffective if the parents used in hybridisation possess unequal distribution of desirable alleles (traits) between the parents (say 70:30 or 80:20). It is because the frequency of the desirable genotypes in early or later segregating generations is very low. Searching for the best 80:20 in a population containing a full of an array of 50:50 types at a very low frequency occurrence of these different 80:20 types is perhaps equivalent to searching a pin in a haystack. Very often plant breeders fail to find potential transgressive segregants in the segregating generations and it is therefore important to understand reasons for the failures which often depends on how the segregating populations are handled during selection and stabilisation process.

The methods of handling segregating generations are well defined when the target genotype is 50:50 type (Pedigree/ Bulk/ Single seed descent method) or an extreme type such as 99:1 or 98:2 or 100:0 (Backcross breeding). It is advised to practice limited backcross breeding when the target genotype is in between these two types, say 70:30 or 80:30. In a single backcross derived populations (BC_1F_2, BC_1F_3... etc.,), 75:25 types are most commonly observed and hence it is easier to expect higher frequency of 70:30, 75:25 or 80:20 types (Fig. 7). Here again, the task of the breeder is simplified to the extent of finding which 75 are seen from the first parent and which 25 are seen from second parent. When majority of segregants are 75:25 types, it becomes easy to pick up the extreme positive transgressive segregants with the required desirable 75 alleles from the first parent and 25 from the second parent.

Thus, selection in a limited backcross-derived population (single backcross) would be successful when the desirable alleles are unequally distributed between the parents (say around 75:25) while selfed breeding methods will be more efficient when parents possess nearly equal distribution of desired alleles between them. The utility of limited backcross approach of breeding has been highlighted earlier by Patil (2007, 2011).

Backcross derived Pedigree/Bulk/Single seed descent method

It is important to note that limited backcrossing just refers to creation of base BC_1F_2 population with higher frequency of the target genotype. Once such base population is created, the breeder has many options of initiating artificial selection right from the BC_1F_2 generation. This modified approach is similar to pedigree method of breeding and can be continued by implementing selection schemes described for F_3, F_4 etc., in the corresponding BC_1F_2, BC_1F_3 etc. generations, respectively. If the considerable heterozygosity existing in a backcross derived population is considered as a disadvantage, then the artificial selection can be delayed up to BC_1F_5 when the proportion of homozygous plants is increased substantially. Either bulk method of advancing or SSD approach can be followed through these early segregating generations. There is a need for research on comparing efficiency of following Pedigree/SSD/Bulk methods in handling segregating generations derived through limited backcross breeding

Following the approach of determining the target genotype involves comparison of backcross Populations to P_1 (BC_1) and P_2 (BC_2) with F_2 population. The segregation pattern observed in BC_1, BC_2 and F_2 populations shows differences in prominent types namely 75:25, 50:50 and 25:75 types in them. This becomes the genetic basis for understanding the differences in means of these populations. If the target genotype set by the pair of parents involved in hybridisation matches with the prominent segregant type observed in a generation, the mean performance of that population will be higher than the mean of the remaining two populations. As per this basis of inheritance, any of the following three situations can be observed in an evaluation study where these three segregating populations are compared.

Table 2: Comparison of Means of Segregating Populations

Situation I: *Decreasing order of performance being $BC_1 > F_2 > BC_2$*

Population	Highest Proportion of alleles of parents		Method of breeding	Result of selection in the population
	P_1	P_2		
BC_1 (F_1 X P_1)	75	25	Limited backcross breeding (with P_1)	More rewarding
F_2 (Selfing F_1)	50	50	Pedigree/Bulk/SSD	Less rewarding
BC_2 (F_1 X P_2)	25	75	Limited backcross breeding (with P_2)	Least rewarding

Here the P_1 parent has higher proportion of desired alleles contributing to higher productivity, and as a result of this, the decreasing order of performance of populations will be $BC_1 > F_2 > BC_2$. In such a case, selfed generations of BC_1 populations can be subjected to selection either by following pedigree or bulk or single seed decent methods of breeding.

Situation 2: *Decreasing order of performance being $BC_2 > F_2 > BC_1$*

Population	Highest Proportion of alleles of parents		Method of breeding	Result of selection in population
	P_1	P_2		
BC_1 (F_1 X P_1)	75	25	Limited backcross breeding (with P1)	Least rewarding
F_2	50	50	Pedigree/Bulk/SSD	Less rewarding
BC_2 (F_1 X P_2)	25	75	Limited backcross breeding (with P2)	More rewarding

Here the P_2 parent has higher proportion of desired alleles contributing to higher productivity because of which these populations reveal a decreasing order of performance of $BC_1 > F_2 > BC_2$. In this situation, selfed generation of BC_2 populations can be subjected to selection either by following pedigree or bulk or single seed decent method of breeding.

Situation 3: *Decreasing order of performance being $F_2 > BC_1$ or BC_2.*

Population	Highest Proportion of alleles of parents		Result of selection in population
	P_1	P_2	
F_2	50	50	More rewarding
BC_1	75	25	Less rewarding
BC_2	25	75	

Here P1 and P2 parents have an equal proportion of desired alleles contributing to higher productivity of F2>BC1>BC2 or F2>BC2>BC1. In this situation advancing selfed F2 and later segregating generations can be subjected to selection either by following pedigree or bulk or SSD method of breeding.

Studies on determining the 'target genotype'

To test this concept, a study was conducted to understand the differences in performance of the three segregating populations and their subsequent selfed generations in cotton at Dharwad, India. Among the three crosses handled in RAH100 X SAM 4, BC_1 population revealed higher mean seed cotton yield/plant and this trend of superiority was continued in the next selfed generation, confirming that the trend of observed superiority occurring in this population did continue in the next selfed generation (Table 3). Many transgressive segregants were noticed in this population. As per the expectations of the genetic basis explained earlier, RAH 100 has many important yield contributing characters including boll number because of which BC_1 and the subsequent selfed population revealed improved performance as compared to the remaining two populations. In a cross of DSC 7 X RAH 53 involving a compact parent and a stay green type, F_2 population was significantly superior to B_2 indicating that the parents were complimenting each other for the yield influencing loci due to which F_2 revealed superior performance and this trend of superiority continued in F_3 generation. In RAH 111 X RAH 16, BC_2 population revealed higher population mean as compared to the other two populations and this trend continued in next selfed generation. The number of transgressive segregants observed in these populations in general matched with the higher mean seen in the populations. Thus, different combinations of parents included in the study differed with respect to the pattern of complementation of genes for which they differ.

The parental genotypes used in the study were evaluated for an array of yield attributing characters including physiological characters related to photosynthesis. An overall assessment of parents for these many important traits gave insight on complementation pattern of desirable traits (alleles) between the elite genotypes chosen. This assessment matched with the previously mentioned conclusion derived from comparing the three segregating generations — F_2, BC_1 and BC_2.

In another study, different populations with gradation of allelic contributions from the two parents were developed. This study involved two intra-*hirsutum* crosses involving four varietal lines 1) DRGR 2572 x M 5 and 2) DRCR 4 X DSMR 10. The nine generations (P_1, P_2, F_1, BC_1, BC_2, F_2, BC_1F_2, BC_2F_2, F_3) of both crosses were sown and evaluated in kharif 2015. The data on these generations confirmed that the trend of superior performance and gradations in performance is confirmed in succeeding selfed generations. Further in an effort to exploit the variability released in these populations top 30% of the best plants were selected from the BC_1, BC_2, F_2, BC_1F_2, BC_2F_2, F_3 generations and All the nine generations (P_1, P_2, F_2, BC_1F_2, BC_2F_2, F_3, BC_1F_3, BC_2F_3, $F_{4,}$) of both crosses

Table 3. Comparison of Performance of Limited Backcross and Selfed Populations for Seed Cotton Yield (g/plant)

Cross	2007-2008			2009-2010		
I. RAH100 X SAM4	BC_1	F_2	BC_2	BC_1F_2	F_3	BC_2F_2
No. of plants/lines (L)	154	468	198	96 L	287 L	71 L
Mean Seed Cotton Yield(g/pl)	167.1*	144.3	129.8	138.6*	118.5	99.4
II. DSC 7X RAH 53						
No. of plants/lines (L)	166	398	184	78 L	243 L	97 L
Mean Seed Cotton Yield(g/pl)	174.8	189.1*	131.7	142.8	154.9*	139.6
III. RAH 111 X RAH 16						
No. of plants/lines (L)	165	371	200	77 L	216 L	90 L
Mean Seed Cotton Yield(g/pl)	157.7	150.4	169.0	121.4	146.2	164.1*

*Significantly higher.

were sown for evaluation in kharif 2016 (Adarsha, 2017). The superior variability observed in the specific populations responded to selection practiced for high yield. This confirms that the higher mean and variability seen in back cross populations responds to selection and confirms that these modified procedures are effective in achieving higher genetic gain for productivity enhancement.

Genetic basis of exploiting heterotic groups and improving performance of hybrids

Performance of hybrids depends upon genetic diversity and complementation of parents for loci influencing different important traits between parents. In a self-pollinated crop like cotton, in which inbreeding depression is not much of a concern, it is possible to observe *per se* performance by grouping genotypes into heterotic groups before understanding their combining ability pattern. Continuous efforts made on relating hybrid performance in cotton with genetic diversity existing among parents has led to development of heterotic groups in cotton. In general, crosses between robust, bushy types and compact types were found to be heterotic because of complementation between these groups. In addition, the stay green groups of genotypes combined well with high 'relative growth rate' types and also compact types. Several studies were conducted to identify elite combiners of each group and exploit these heterotic groups for improving combining ability by practicing procedures such as reciprocal selection for combining ability (Patil et al 2007, Patil and Patil 2007, Patil et al., 2011, Pranesh and Patil 2014). Initially, variability for favourable dominant alleles was created by using two parents in each group. In recent years, broad based populations are developed based on four parents in each group. Other studies on checking consistency of combining ability over generations was carried out. Though potential crosses were obtained between F_3 lines and the opposite groups testers, performance of crosses between the new double cross F_3 lines of the opposite groups were found to be of more potential. This indicates diversity created between the lines of opposite groups obtained from four parent crosses. Development of populations of opposite groups based on four elite combiner parents appears to be a promising approach in exploiting heterotic groups.

Two different breeders with varied understanding of genetic principles can produce different results even after handling the same population for developing a variety or the same germplasm for developing a hybrid. The innovative approaches followed in the creation of useful variability required for varietal development and modifications in approaches of exploiting this variability have shown encouraging results indicating in general that instead of the conventional varietal development approaches there is a need to incorporate these innovative modifications in breeding methodologies for developing superior varieties and hybrids.

Developing broad-based populations under national programmes

There is a need to develop broad based populations by pooling genes of resistance to diseases like CLCuV, and pests such as jassids and bollworms, which are known for their ability to quickly develop resistance to protection measures used against them. Plant breeding solutions of incorporating such special traits should involve pooling genes for resistance by utilising resistant sources. In addition to wild species, diploid species can also be explored for improvement of these traits. The approaches of introgression through systematically planned, limited backcrossing may become necessary. The size of a backcross population grown/found to be resistant/and used for backcrossing needs to be distinguished. The optimum/minimum population sizes assume astronomical figures and the implications of a limited sample size on the chance of recovering 'recurrent parents constitution' becomes more difficult than what is assumed in consequences of backcrossing.

National/continental programs of exploring useful alleles in wild species through limited introgression

Wild species need to be explored for the hidden valuable genes, alleles for stress tolerance and rare alleles associated with the main productivity traits. It is essential to initiate national programs of developing backcross programs involving wild species and diploid species. While practicing backcrossing, it is essential to understand and distinguish concepts of minimum and optimum population sizes to be grown with plants carrying the desirable target trait and their influence on recovering constitution of the recurrent parent.

Use of induced mutations to supplement and compliment recombinational variability

It is possible to complement hybridisation with induced mutations to generate more useful variability. In addition, induced variability can be helpful in producing rare alleles for biotic and abiotic stress resistance. Apart from inducing mutations for creation of variability in varietal breeding, it can be used to create variability for gene action and multiple alleles helpful in hybrid breeding.

Development of heterotic populations for common use by breeders

The modified procedures of 'reciprocal recurrent' selection which are basically defined for cross pollinated crops are aimed at increasing genetic diversity between opposite populations. The phenomenon of heterosis is basically having the same genetic basis of $H_{F1} = \sum dy^2$ and implications of the formula clearly show that by increasing genetic diversity, it is possible to increase the magnitude of heterosis. A series of these studies on exploiting heterotic

groups have proved that it is possible to transcend the limits of mating system and utilise the procedures originally described for cross-pollinated crops used for improving self-pollinated crops as well.

Options of breeding genetically modified (GM) events

Populations or pools of alleles should be built with active and effective GM events. Apart from deregulated events like Mon531, some new events are being compared and many more are being developed and identified. New events from UAS, Dharwad (cry1Ac event, UASD No. 78), Delhi University (cry1Ac event, Tg2E13), NBRI, Lucknow (Tma12 & Cry1Ec events) are prominent among them. There is a need for identifying the most effective ones among these events to incorporate them into elite varieties and release them for commercial cultivation through fast-track approaches. With appropriate pyramiding or stacking of potential events, it is possible to achieve sustainable cotton production in the years to come. Providing access to these events for the breeders in state agricultural universities (SAUs) can expedite the progress.

Need for special support for conventional breeding research

There is a severe dearth of support for plant breeding research and procedures of varietal improvement and hybrid development. This paper draws the attention of cotton researchers to the recent findings from our plant breeding research and also toward some inferences and plant breeding implications derived from the basic principles of 'population and quantitative genetics'.

Programmes on varietal development and hybrid development that need to be approached by a team of breeders, covering different locations are highlighted above. Some additional joint ventures of teams of scientists which need to be supported with an open mindset are explained. Plant breeders often experience that that innovative ideas of research submitted as research projects generally do not attract the attention of funding agencies and this situation is highlighted in the cartoon (Figure 8). While molecular approaches can complement plant breeding approaches, there is a need for funding agencies as well as government institutions to appreciate the importance of innovative conventional breeding methodologies that may provide a much higher value of returns for investments than the investment made on molecular aspects for plant breeding.

References

Adarsha.,2017 Genetic studies on target genotype and response to selection in cotton. M.Sc(Agri) thesis submitted to University of Agricultural Sciences,Dharwad.p143

Allard R.W., 1960., Principles of Plant breeding, John Wiley and sons, Inc. New York, P.485

Patil, S. S. and Patil, S. A., 2003, Role of improving combining ability in increasing performance of cotton hybrids. Third World Cotton Res. Conference, Cape Town South Africa. 9-13 March 2003.

Patil, S. S., Patil, R. S., Deepak, C. A., Somashekhar, D. and Kolhar, B. C., 2003, Genetic evaluation and comparison of *G. herbaceum* and *G. hirsutum* varieties in different levels of rainfed and intensive management situation. Third World Cotton Res. Conference, 9-13 March, Cape Town, South Africa.,

Patil, S. S. and Patil, S. A., 2003, Use of plant type concept in Cotton breeding. Third World Cotton Res. Conference, Cape Town South Africa. 9-13 March 2003. P- 34.

Patil, S. S., Patil, S. A., Ramakrishna, V., Deepak, H. B. and Mallikarjun, H. B., 2007, Approaches of improving combining ability for enhancing performance of cotton hybrids. Proc Fourth World Cotton Research Conference held at Lubbock, USA, 1818.

Patil, S. S., 2007, Potential of limited backcross breeding in improving cotton. Proc Fourth World Cotton Research Conference held at Lubbock, USA, 1926.

Patil, S. S., Ramakrishna, V., Maralappanavar, M. S., Swathi, P., Ranganatha, H. M., Kencharaddi, H. G. and Babu, D. H., 2011, Deploying reciprocal selection for combining ability for improving performance of hybrids in cotton (*Gossypium hirsutum* L.). *Indian J. Genet*., 71(2): 180-184.

Patil, S. S., 2011, Importance of teaching concepts of population genetics in effective understanding of plant breeding. *Indian J. Genet*., 71(2): 106-114.

Patil, S. S., 2011, Target genotype based choice of breeding methods in cotton. World Cotton Research Conference – 5, November 5 – 10, Mumbai, India.

Pranesh, K. J. and Patil, S. S., 2014, Identification of transgressive segregants for combining ability in early generation of cotton. Green Farming, 5(5): 734-737.

Patil, S. S., 2014, Conventional breeding of cotton needs to change. ICAC Recorder, 3:13-20.

Vijayalaxmi, Udikeri, G., Narayanan, K., Nachne, R. P., Patil, S. S. and Kulkarni, V. N., 2011, Trends in cotton fibre quality improvement in India. *Cotton Res. J.,* 25: 84-92.

Vinayak Edke.,. 2016 Genetic studies on utilising heterotic groups for varietal improvement in cotton M.Sc(Agri) thesis submitted to University of Agricultural Sciences, Dharwad, p143.

The ICAC Recorder

Mini-Review

Genome-wide Quantitative Trait Loci Reveal the Genetic Basis of Cotton Fibre Quality and Yield-related Traits

Zhen Zhang[1,2†],Junwen Li[1,3†], Muhammad Jamshed[1,5†+], Yuzhen Shi[1†], AiyingLiu[1],JuwuGong[1], ShufangWang[1],Jianhong Zhang[1], FudingSun[1], Fei Jia[1], QunGe[1], LiqiangFan[1], ZhibinZhang[1], JingtaoPan[1], SenmiaoFan[1], YanlingWang[1], QuanweiLu[4], RuixianLiu[1], XiaoyingDeng[1], Xianyan Zou[1], Xiao Jiang[1], Ping Liu[1],PengtaoLi[4], Muhammad Sajid Iqbal[1], ChuanyunZhang[7], Juan Zou[7], Hong Chen[8], Qin Tian[8], XinheJia[9], BaoqinWang[9],NijiangAi[10], GuoliFeng[10],YumeiWang[11], Mei Hong[12], ShilinLi[12], WenmingLian[13], Bo Wu[13],Jinping Hua[6], ChaojunZhang[1], JinyongHuang[3], Aixia Xu[2*], Haihong Shang[1,3*], Wankui Gong[1,4*] and Youlu Yuan[1,4*]

Keywords: upland cotton, consensus genetic map, fibre quality, fibre yield, QTL clusters, genetic correlation, gene expression level

Summary

Cotton is widely cultivated globally because it provides natural fibre for the textile industry and human use. To identify quantitative trait loci (QTLs)/genes associated with fibre quality and yield, a recombinant inbred line (RIL) population was developed in upland cotton. A consensus map covering the whole genome was constructed with three types of markers (8295 markers, 5197.17 centimorgans (cM)). Six fibre yield and quality traits were evaluated in 17 environments, and 983 QTLs were identified, 198 of which were stable and mainly distributed on chromosomes 4, 6, 7, 13, 21, and 25. Thirty-seven QTL clusters were identified, in which 92.8% of paired traits with significant medium or high positive correlations had the same QTL additive effect directions, and all of the paired traits with significant medium or high negative correlations had opposite additive effect directions. Six important QTL clusters that included both fibre quality and yield traits were identified with opposite additive effect directions, and those on chromosome 13 (*qClu-chr13-2*) could increase fibre quality but reduce yield; this result was validated in a natural population using three markers. These data could provide information about the genetic basis of cotton fibre quality and yield and help cotton breeders to improve fibre quality and yield simultaneously.

Introduction

Among the four cultivated *Gossypium* species, *G. hirsutum* (upland cotton) accounts for 90% to 95% of the total cotton fibre production because of its wide adaptability and attractive fibre qualities (Percival *et al*., 1999; Rong *et al*., 2004; Park *et al*., 2005; Lacape *et al*., 2009). In conventional breeding procedures, which use sexual hybridisation, parents are selected based on their desired phenotypic performance for the target traits. For complex traits such as cotton fibre quality and yield, however, it becomes chal-

[1] State Key Laboratory of Cotton Biology: Key Laboratory of Biological and Genetic Breeding of Cotton, The Ministry of Agriculture; Institute of Cotton Research, Chinese Academy of Agricultural Sciences; Anyang, Henan, 455000 China
[2] State Key Laboratory of Crop Stress Biology for Arid Areas, College of Agronomy, Northwest A&F University, Yangling, China
[3] Zhengzhou Research Base, State Key Laboratory of Cotton Biology, Zhengzhou University
[4] Research Base, Anyang Institute of Technology, State Key Laboratory of Cotton Biology, Anyang, Henan, 455000 China
[5] Department of Biochemistry University of Agriculture, Faisalabad, Pakistan
[6] College of Agronomy and Biotechnology, Key Laboratory of Crop Heterosis and Utilization of The Ministry of Education, Beijing Key Laboratory of Crop Genetic Improvement, China Agricultural University, Beijing 100193, China
[7] Shandong Cotton Research Center, Jinan,250100, China
[8] Cotton Research Institute, Xinjiang Academy of Agricultural and Reclamation Science, 832000, China
[9] Zhengzhou Research Institute of Agricultural and Forestry Sciences, Zhengzhou 450000, China
[10] Shihezi Agricultural Science & Technology Research Center, Shihezi 832000, China
[11] Institute of Cash Crops, Hubei Academy of Agricultural Science, Wuhan 430064 China
[12] Institute of Agriculture, Bayingolin Mongol Autonomous Prefecture, Korla, Xinjiang 841000, China
[13] Institute of Agricultural Science and Technology, Agricultural Production Division 1, Xinjiang Production and Construction Corps, Alar Xinjiang 843300, China
†: These authors made equal contributions
*: Corresponding authors: Youlu Yuan (yuanyoulu@caas.cn),Wankui Gong (wkgong@aliyun.com), or Haihong Shang (shanghaihong@caas.cn), Aixia Xu (799170439@qq.com)

lenging to further improve one without compromising the other (Shen *et al*., 2005; Qin *et al*., 2008; Yu *et al*., 2013; Ning *et al.*, 2014; Cao *et al.*, 2015; Li *et al.*, 2016a). Therefore, uncovering the genetic basis of the fibre quality and yield traits in upland cotton is becoming increasingly necessary for the simultaneous improvement of fibre quality and yield in the future (Sun e*t al.*, 2012; Ning *et al.*, 2014).

Sequencing of the genomes of *G. raimondii* (Paterson *et al*., 2012; Wang *et al*., 2012), *G. arboreum* (Li *et al*., 2014; Du *et al*., 2018), *G. barbadense* (Liu *et al*., 2015; Yuan *et al*., 2015), and *G. hirsutum* (Li *et al.*, 2015; Zhang *et al.*, 2015b;Wang *et al*., 2019) has provided new platforms for linkage-map-based QTL identification. To explain the genetic basis of cotton fibre quality and yield-related traits, many genetic linkage maps have been constructed, including interspecific maps mainly between *G. barbadense* and *G. hirsutum* (Chee *et al.*, 2005a and 2005b; He *et al.*, 2007; Lacape *et al.*, 2010; Yang *et al.*, 2015;Zhai*et al.*, 2016;Wang *et al.*, 2016;Song *et al.*, 2017; Chandnani *et al.*, 2018)and intra-*hirsutum* maps (Sun *et al*., 2012; Zhang *et al*., 2012, 2016 and 2017; Ning *et al.*, 2014, Tan *et al*., 2014, Tang *et al.*, 2015, Shang *et al.*, 2015;Liu *et al*., 2*017);* QTLs have been identified based on these maps. Restriction fragment length polymorphism (RFLP) markers were used to construct the initial genetic map in cotton (Peterson *et al.*, 2003; Ualloa *et al.*,2002 and 2005; Rong *et al.*, 2004), followed by simple sequence repeat (SSR) markers (Sun *et al*., 2012; Zhang *et al*., 2012 and 2015; Tan *et al*., 2014; Tang *et al.*, 2015; Yang *et al.*, 2016; Liu *et al.*, 2017) and single nucleotide polymorphism (SNP) markers (Hulse-Kemp *et al*., 2015; Wang *et al*., 2015a and 2015b; Li *et al*., 2016a; Zhang *et al*., 2016 and 2017; Qi *et al*., 2017; Diouf *et al*., 2018). A QTL is a chromosome region that might cumulatively contribute to phenotypic variation; each QTL has a varying degree of efficacy for a specific trait (Ulloa *et al*., 2000; Peterson *et al*., 2003; Rong *et al*., 2004; Shen *et al*., 2005 and 2007; Zhang *et al*., 2012; Tan *et al*., 2014; Wang *et al*., 2015b and 2015c). QTLs for fibre quality and yield traits have contributed to cotton breeding through marker-assisted selection, fine mapping, functional characterisation of genes, and gene cloning. However, the maps used did not cover the whole genome and the environments used to assess the phenotypes were insufficient (Hulse-Kemp *et al*., 2015; Tang *et al*., 2015; Wang *et al.*, 2015a and 2015b, Li *et al*., 2016a; Liu *et al*., 217; Qi *et al*., 2017; Diouf *et al.*, 2018; Jia *et al*., 2018). Thus, many QTLs for fibre quality and yield traits remain to be identified and characterised across the upland cotton genome. It remains necessary to dissect the genetic basis of traits related to fibre quality and yield by identifying QTLs using a high-density genetic linkage map with coverage of the entire upland cotton genome.

In this study, we developed a population of intraspecific RILs in upland cotton consisting of 196 lines with the parents 0-153 and sGK9708. The fibre quality and yield traits were evaluated across 22 environments. A whole-genome-based high-density genetic linkage map was constructed by integrating five previously reported basic linkage maps generated using SSR (Sun *et al*., 2012; Jamshed *et al*., 2016), SSR and SNP-BSA (bulked segregation analysis) (Zhang *et al*., 2015c), SNP-specific-locus amplified fragment sequencing (SLAF) (Zhang *et al*., 2016b), and SNP-chip (Zhang *et al*., 2017b). QTLs for fibre quality and yield traits were identified in 17 of the 22 environments. QTL clusters were also identified to determine the genetic basis of the correlation between fibre quality and fibre yield.

This study demonstrates the integration of various linkage maps with different marker types to construct a high-density genetic map covering the whole genome. This technique could be a powerful complementary strategy to dissect the genetic basis of complex traits and their genetic correlations in cotton. In addition to providing new insights into the genetic basis of the traits related to fibre quality and yield and their relationships, this study also provided information about improving cotton fibre quality and yield simultaneously.

Results

Phenotypic variation and correlation of traits relating to fibre quality and yield

A RIL population developed between the two parents 0-153 and sGK9708 was planted from 2007 to 2015 across 22 environments. The phenotypic variation ranges of fibre strength (FS), fibre length (FL), and micronaire (FM) for the parents are summarised in Figure 1. All the traits

Table 1. The Phenotypic Variation Range of Fibre Strength (FS), Fibre Length (FL) and Micronaire (FM)

Year	parents	Fiber strength	Fiber length	Micronaire
		cN/tex	mm	
2003 Anyang	sGK9708	25.00±0.31	27.22±0.14	3.94±0.11
	0-153	32.31±0.72	28.55±0.38	4.33±0.19
2004 Anyang	sGK9708	25.00±0.21	28.68±0.11	4.39±0.11
	0-153	35.88±0.44	29.04±0.20	4.29±0.07
2007 Anyang	sGK9708	25.65±1.20	27.44±0.01	4.68±0.18
	0-153	33.75±1.06	31.14±0.17	4.50±0.09
2008 Anyang	sGK9708	25.60±0.71	28.18±0.08	5.35±0.12
	0-153	32.60±0.85	31.24±0.06	5.00±0.13
2008 Quzhou	sGK9708	26.70±1.56	27.54±0.87	5.37±0.03
	0-153	34.50±0.57	30.21±0.51	4.28±0.11
2008 Linqing	sGK9708	24.20±0.28	26.67±0.11	5.19±0.14
	0-153	33.40±1.27	30.00±0.68	4.54±0.23

showed approximately normal distributions with an absolute skewness value of less than one and were characterised by transgressive segregation with respect to parental performance.

The heritability for FS, FL, FM, BW, and LP were calculated across 21 environments, excluding 2013 in Anyang; the heritability for SI was calculated across 20 environments, excluding 2012 in Zhengzhou and 2013 in Anyang. The heritability for all traits was more than 50%, except in 2015 Shihezi (49.80%) and 2014 Alaer (31.20%), which indicates that these six traits had high heritability.

To establish correlations between different traits in the same environment, three categories of correlations were observed among the six traits. In the first category, the trait pairs (FS and FL, FS and SI, FL and SI, FM and BW, and BW and SI) showed significant medium or high positive correlations in most of the environments; in the second category, the trait pairs (FS and FM, FS and LP, FL and FM, FL and LP, and LP and SI) showed significant medium or high negative correlations in most of the environments; and in the third category, the trait-pairs (FM and LP, FS and BW, FL and BW, FM and SI, BW and LP) showed no or weak positive or negative correlations in most of the environments (Figure 1).

The correlations between the traits differed significantly from those in the other environments In 2012 in Zhengzhou, in 2014 in Kuerle, and in 2015 in Kuerle. Some traits did not show any significant positive correlations with those in the other environments in 2011 in Anyang, in 2011 in Zhengzhou and in 2012 in Zhengzhou. These results indicated that these five environments were distorted; thus, the other 17 environments were used for QTL identification.

Genotyping for constructing the genetic linkage map

To ensure the quality of the genetic consensus map, all three types of markers (SSR, SLAF-SNP, and chip-SNP) were first filtered. For the SLAF-SNP markers, the processing and the results of filtering the markers are shown in Zhang's reports (Zhang *et al.*, 2016). For the chip-SNP markers, 6989 SNPs were shown to be polymorphic between the parents, 4863 were located on the genome of upland cotton, 2734 were genotyped as AA × BB with a missing rate of less than 40%, and 2315 showed nonsignificant segregation distortion with *P* values of more than 0.001 and could therefore be used to construct the consensus genetic map (Zhang *et al*., 2017b). For the SSR markers, 398 markers with 458 loci from Sun's and Jamshed's research could be located in the upland cotton genome and used to construct the consensus genetic map; these markers were developed by other organisations (Sun *et al*., 2012; Jamshed *et al*., 2016).

Construction of the consensus genetic linkage map

After combining the filtered markers, the consensus genetic linkage map consisted of 8295 markers (458 SSRs, 5521 SLAF-SNPs, and 2316 chip-SNPs) and spanned 5197.17

Table 2. Detailed Information on the Consensus Genetic map for the RIL Population

No. Chr	Marker Num	Total Genetic Distance (cM)	Average Genetic Distance (cM)	Largest Gap	No. SLAF-SNP Markers	No. chip-SNP Markers	No. SSR Markers	Nmuber of Gap(>5cM)	Number of SDMs (0.00 1<=P<0.05)	Percentage of SDMs (%)	chi-square	P_value	SDR Region
chr01	414	181.81	0.44	6.77	297	108	9	1	119	28.74	2.84	0.21	3
chr02	238	151.99	0.64	11.18	180	54	4	6	21	8.82	1.32	0.45	2
chr03	329	191.94	0.59	16.29	218	100	11	5	68	20.67	2.35	0.48	3
chr04	646	208.85	0.32	15.12	574	60	12	7	8	1.24	0.94	0.47	1
chr05	571	221.34	0.39	8.13	434	119	18	5	196	34.33	2.87	0.29	5
chr06	181	207.63	1.15	16.98	101	42	38	3	76	41.99	6.80	0.35	4
chr07	422	168.99	0.4	4.76	318	97	7	0	29	6.87	1.13	0.45	2
chr08	78	85.3	1.11	4.79	56	21	1	0	20	25.64	7.01	0.31	1
chr09	376	193.16	0.52	14.15	274	91	11	5	90	23.94	2.60	0.40	7
chr10	218	199.12	0.92	9.90	133	68	17	11	74	33.94	9.36	0.24	2
chr11	120	154.86	1.3	5.39	88	28	4	1	5	4.17	1.55	0.33	1
chr12	368	212.73	0.58	4.76	273	89	6	0	89	24.18	2.61	0.24	5
chr13	766	207.22	0.27	5.07	604	151	11	1	64	8.36	1.67	0.36	6
chr14	727	234.63	0.32	2.98	408	305	14	0	477	65.61	7.32	0.15	2
chr15	85	226.61	2.7	13.20	29	27	29	18	8	9.41	1.49	0.38	1
chr16	620	245.76	0.4	16.17	399	175	46	8	202	32.58	5.37	0.28	5
chr17	184	199.41	1.09	16.59	102	60	22	10	11	5.98	1.01	0.51	1
chr18	286	236.88	0.83	23.66	172	96	18	15	52	18.18	2.10	0.31	4
chr19	180	219.65	1.23	13.95	109	56	15	14	50	27.78	2.67	0.35	5
chr20	150	226.79	1.52	13.51	60	62	28	12	43	28.67	6.03	0.34	4
chr21	271	221.93	0.82	5.44	174	89	8	2	8	2.95	0.90	0.53	2
chr22	149	152.86	1.03	4.76	75	64	10	0	2	1.34	0.54	0.65	0
chr23	302	244.88	0.81	13.29	142	131	29	7	109	36.09	7.95	0.22	6
chr24	109	132.17	1.22	3.90	60	46	3	0	16	14.68	1.82	0.43	1
chr25	353	230.04	0.65	24.36	166	121	66	7	249	70.54	12.16	0.10	8
chr26	152	240.62	1.59	16.07	75	56	21	12	22	14.47	1.86	0.42	3

Table 3. The Number of QTLs and Stable QTLs for the Six Fibre Quality and Yield Traits in 17 Environments

No.Chr	FS		FL		FM		BW		LP		SI		Total	
	Total	Stable	Total	Stable	Total	Stable	Total	Stable	Total	Stable	Total	Stable	Total	Stable
Chr01	1	0	12	4	7	1	8	4	2	1	4	0	34	10
Chr02	0	0	3	0	3	2	6	0	4	1	5	1	21	4
Chr03	4	0	4	0	6	0	5	1	8	3	6	0	33	4
Chr04	12	5	13	8	14	3	8	0	8	3	13	4	68	23
Chr05	7	0	7	1	8	2	11	2	5	0	5	1	43	6
Chr06	10	3	4	2	8	2	18	3	6	2	5	1	51	13
Chr07	14	3	16	3	8	2	13	4	9	2	22	7	82	21
Chr08	4	2	2	1	6	1	2	0	5	0	2	0	21	4
Chr09	4	0	1	0	1	1	10	3	5	0	5	0	26	4
Chr10	1	0	5	0	12	1	16	2	7	0	7	1	48	4
Chr11	7	1	10	2	5	2	4	1	3	0	5	2	34	8
Chr12	1	0	2	0	5	0	12	1	2	0	1	0	23	1
Chr13	8	3	11	3	11	3	16	4	11	3	7	0	64	16
Chr14	12	1	8	2	12	0	8	0	9	3	8	0	57	6
Chr15	3	0	8	0	4	2	5	1	2	0	1	0	23	3
Chr16	8	3	4	1	8	1	8	3	7	1	8	0	43	9
Chr17	1	0	1	0	3	3	3	1	5	0	5	1	18	5
Chr18	5	1	5	0	7	1	8	0	5	0	1	0	31	2
Chr19	6	1	6	1	6	0	5	0	5	0	9	2	37	4
Chr20	3	1	4	1	3	0	4	0	6	2	3	0	23	4
Chr21	8	2	6	4	6	3	5	1	6	1	8	1	39	12
Chr22	6	2	3	0	6	0	6	1	3	0	8	1	32	4
Chr23	3	0	2	0	4	1	5	1	1	0	3	0	18	2
Chr24	0	0	1	0	2	0	5	1	11	2	5	1	24	4
Chr25	6	5	10	2	13	1	16	8	14	2	12	3	71	21
Chr26	4	0	1	0	2	0	4	1	4	2	4	1	19	4
Total	138	33	149	35	170	32	211	43	153	28	162	27	983	198

cM, with 2384.94 cM for the At sub-genome and 2812.23 cM for the Dt sub-genome. The average interval between adjacent markers was 0.88 cM. The gaps (more than 10cM) in the integrated genetic linkage map were mainly found on D_t-sub-genome chromosomes, with five on chromosome15 and four onchromosome26. The largest map gap was found on chromosome 25 (24.36 cM). The segregation distortion of the markers (SDM) ($0.001<P\text{value}<0.05$) in the integrated linkage map remained at almost the same level as in the component maps. Chromosome14 (447) and chromosome 25 (249) were the two chromosomes with the highest levels of SDM. The marker density in the genetic and physical maps along the whole genome indicated high coverage, although the markers were unevenly distributed throughout the upland cotton genome, and there were large variations across chromosome regions. For the collinearity analysis, most of the SNP loci on the linkage map were in the same order as those on the corresponding chromosomes of the physical map of the upland cotton genome. These results demonstrated that the current linkage map could provide an effective and reliable tool for QTL mapping (Table 2).

QTL identification for fibre quality and yield

Based on the consensus genetic map and the phenotype data across 17 environments, 983 QTLs (548 on A_t, 435 on D_t) for fibre quality and yield were identified. Among them, 198 were stable (118 on A_t and 80 on Dt) in at least three environments, and 53.53% of stable QTLs were mainly distributed on chromosomes 4, 6, 7, 13, 21, and 25, especially the three-environment-stable QTLs (Table 3 and Figure 1).

There were 138 QTLs for FS, 149 for FL, 170 for FM, 211 for BW, 153 for LP and 162 for SI. Among them, 33 (A_t was 17, D_t was 16), 35(A_t was 24, D_t was 11), 32 (A_t was 20, D_t was 12), 43 (A_t was 25, D_t was 18), 28 (A_t was 15, D_t was 13), and 27 (A_t was 17, D_t was 10) QTLs were stable, respectively. These stable QTLs were distributed on all 26 chromosomes of cotton, mainly on chromosomes 4, 6, 7, 11, 13, 16, 20, 21 and 25. Detailed information about the confidence intervals, observed phenotypic variance (PV) values, and additive effects of each QTL is shown in Table 3, and Figure 1.

QTL cluster identification

Among the QTLs identified for fibre quality and yield, there were QTLs for different traits that shared the same confidence intervals and were considered to constitute QTL clusters. In this research, 37 QTL clusters were identified (23 on A_t and 14 on D_t). All these QTL clusters were distributed on 17 chromosomes (chromosomes2, 4, 5, 6, 7, 8, 11, 13, 15, 16, 17, 19, 20, 21, 22, 25, and 26). Six chromosomes (chromosomes 4, 6, 7,13, 21, and 25) harboured 25 QTL clusters, accounting for 67.6% of the total clusters

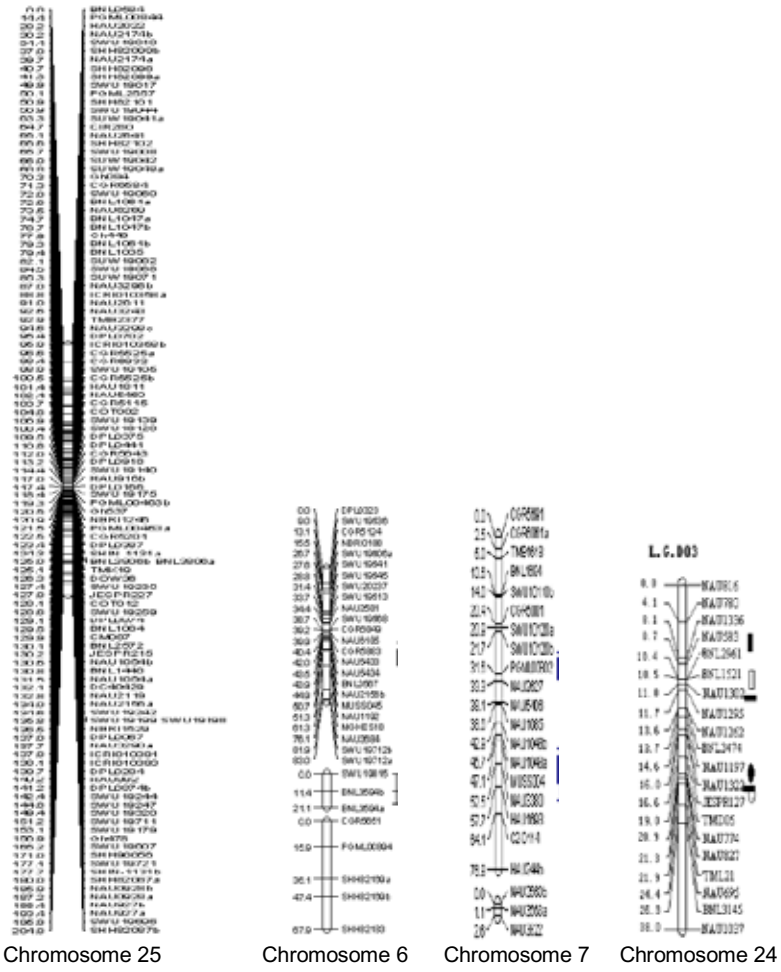


Figure 1. QTL Mapping of Chromosomes 6, 7, 24 and 25 (Jamshed et al. BMC Genomics, 2016. 17:197)

identified, and could be considered important chromosomes in this study. Among these QTL clusters, there were six QTL clusters that harboured QTLs for FS or FL and BW or LP. The 0-153 allele increased FS and FL but decreased LP or BW in the *qClu-chr16-1*, *qClu-chr20-1*, *qClu-chr6-2*,*qClu-chr25-2*, and *qClu-chr13-2*cluster; it also increased LP but decreased FS in *qClu-chr21-3*.

Discussion

Coverage and saturation of the consensus genetic map

Based on the reference genome of upland cotton, the total genetic distance of the genetic map should be between 4500cMand 5500cM without visible gaps (Wang *et al.*, 2012; Peterson *et al.*, 2012; Li *et al.*, 2013 and 2015; Zhang *et al.*, 2015b, Du *et al.*, 2018). In previous studies, several high-density genetic maps were constructed using different markers with different populations; these maps, however, included genetic gaps of more than 20cM (Tan *et al.*, 2014; Hulse-Kemp *et al.*, 2015; Liu *et al.*, 2015 and 2017; Tang *et al.*, 2015;Wang *et al.*, 2015a and 2015b, Li *et al.*, 2016a; Zhang *et al.*, 2016 and 2017; Qi *et al.*, 2017; Diouf *et al.*, 2018; Jia *et al.*, 2018). In this study, we used three types of markers to construct a consensus genetic map that could fill some of the gaps in the previously generated maps. This consensus genetic map covered the whole genome of upland cotton with high saturation and is therefore a valuable tool for QTL and candidate gene identification, functional characterisation, and pyramiding breeding across the whole genome.

QTL numbers compared with previous reports

Because quantitative traits are influenced by the environment, some QTLs could be identified in several environments, while others could be identified only in specific environments; QTLs that could be identified in multiple environments or generations are considered stable QTLs (Sun et al., 2012) and may be of use for marker-assisted breeding and gene cloning. As cotton fibre quality and yield traits are quantitative traits, they are controlled by multiple loci/genes. In previous studies, no more than 16 stable QTLs were identified within a population for each of the fibre quality or yield traits, i.e., the maximum of 16 was found for FS, seven for FL, five for FM, 16 for BW, seven for LP, and four for SI (Tan *et al.*, 2014 and 2018;Hulse-Kemp *et al.*, 2015; Tang *et al.*, 2015;Wang *et al.*, 2015a and 2015b; Li *et al.*, 2016a; Yang *et al.*, 2016; Liu *et al.*, 2017; Qi *et al.*, 2017; Diouf et *al.*, 2018). In this study, we used a consensus genetic map with high coverage and saturation and identified more than 25stable QTLs across 17 environments. Therefore, this study identified the greatest number of stable QTLs for fibre quality and yield. These results could provide more information about the genetic mechanisms underlying cotton fibre development and provide important loci related to the improvement of both cotton fibre quality and yield.

Congruence with previously reported QTLs

To determine whether the QTLs in our study were novel or had been previously identified, we compared our results with those from the cotton QTL database based on their physical confidence intervals and the reports of previous genome-wide association studies (GWAS). A total of 43QTLs shared the same or overlapping confidence intervals with QTLs identified in previous studies (Rong *et al.*, 2007; Lacape *et al.*, 2010; Said *et al.*, 2013, 2015a and

2015b; Fang *et al*., 2014).Of these, 22 were in the cotton QTL database, and 25were identified in previous GWAS (Fang *et al.*, 2017; Huang *et al.*, 2017; Sun *et al.*, 2017; Wang *et al.*, 2017a; Ma *et al*., 2018).

Overall, of the 198 stable QTLs identified in our study, 43 have been identified in previous studies (seven for FS, seven for FL, 11for FM, four for BW, ten for LP, and four for SI.) The others (78.3% of 198) were newly identified (26 for FS,28 for FL,21 for FM, 39 for BW, 18 for LP, and 23 for SI) and could provide more information about the mechanisms of cotton fibre development and yield formation.

Congruence with previously reported QTL clusters

In previous studies, many QTL clusters were reported and are available in the cotton QTL database (Rong *et al*., 2007; Lacape *et al*., 2010; Said *et al.*, 2013, 2015a and 2015b; Fang *et al*., 2014). In this database, all the chromosomes harboured QTL clusters for no fewer than two traits among FS, FL, FM, BW, LP, and SI. Among the 37 QTL clusters identified in our study, there were six QTL clusters (*qClu-chr6-2*, *qClu-chr7-1*, *qClu-chr13-1*, *qClu-chr13-2*, *qClu-chr17-1*, and *qClu-chr19-1*) that shared the same or overlapping confidence intervals with those in the database; the other 31 clusters were newly identified in our research.

Three markers, Marker 8331, Marker 8332 and Marker 8333 related to *qClu-chr13-2*, were selected to genotype a natural population with 278 breeding parents (Figure 2). For all three markers, the phenotypes of the materials with markers and without markers showed significant differences, with *P* values less than 0.05. The FS increased from 9.77 cN/tex to 10.61 cN/tex, and the FL increased from 6.40mm to 4.21mm with the markers. However, the BW decreased from 2.19g to 2.70g, and the LP decreased from 3.17% to 4.20% with the markers. These results could provide information about the mechanism underlying the negative correlation between yield traits and fibre quality traits and act as a reference for improving the yield and fibre quality of cotton.

QTL clusters and genetic correlation between fibre quality and fibre yield

Some previous studies in cotton have addressed the correlation between traits and QTL clusters but did not combine these results for further analysis (Tan *et al*., 2018; Li *et al*., 2016a; Liu *et al*., 2017; Diouf *et al*., 2018). In these studies, the number of QTL clusters ranged from 2 to 18, and the paired-trait QTL clusters ranged from 2 to 27. Through detailed analysis of these results, we found that when paired traits had a medium or high significant positive correlation, the QTL additive effect directions for these two traits were the same (positive or negative) in most QTL clusters (Liu *et al*., 2017; Diouf *et al.*, 2018;Tan *et al*., 2018); when paired traits had a medium or high significant negative correlation, the QTL additive effect directions for these two traits were opposite (positive and negative) in most of the QTL clusters (Liu *et al*., 2017); and when paired traits had no or only a weakly significant correlation, the QTL additive effect directions for these two traits were sometimes the same and sometimes opposite (Liu *et al*., 2017; Tan *et al*., 2018).

In our research, a total of 59 paired-trait QTL clusters were identified for the fibre yield and quality traits; 28 (47.5%) showed the same QTL additive effect direction (positive or negative), and 31 (52.5%) showed the opposite QTL additive effect directions (positive and negative).

Five paired traits (FS and FL, FS and SI, FL and SI, FM and BW, and BW and SI) showed significant medium or high positive correlations in most environments; 27 paired-trait QTL clusters were identified for these traits. Except for two of the paired-trait QTL clusters between FM and BW and SI and BW, the other 25 paired-trait QTL clusters (92.8%) showed the same QTL additive effect directions (positive or negative), especially the 14 QTL clusters identified between FS and FL.

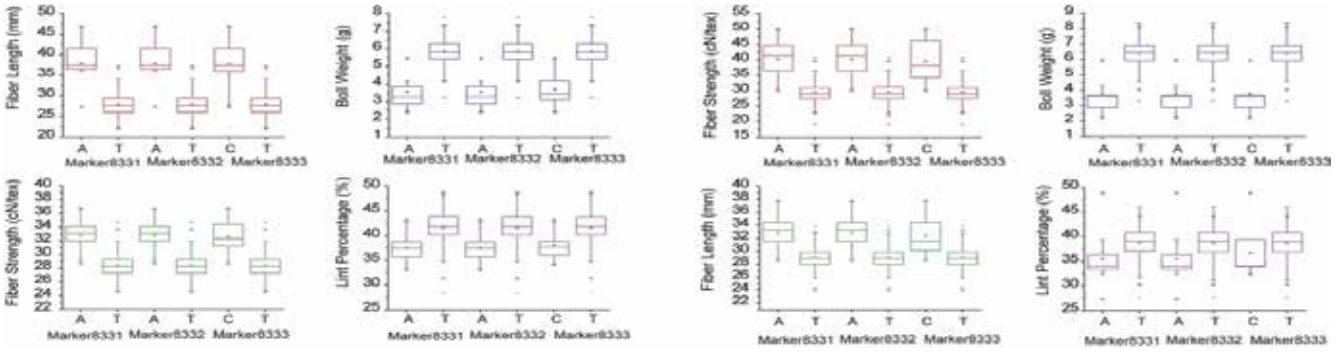


Figure 2. The three makers located in the confidence interval of qClu-chr13-2 could increase FS from 9.77 cN/tex to 10.61 cN/tex; FL from 4.21mm to 6.40 mm with the markers. While decrease BW from 2.19 g to 2.70 g, LP from 3.17% to 4.20% (Zhang zhen *et al. Plant Biotechnology Journa*l, 2019)

Five paired traits (FS and FM, FS and LP, FL and FM, FL and LP, and LP and SI) showed significant medium or high negative correlations in most environments, and 16 paired-trait QTL clusters were identified for these traits, all of which (100%) showed opposite QTL additive effect directions (positive and negative).

Five paired traits (FM and LP, FS and BW, FL and BW, FM and SI, and BW and LP) showed no or weak positive or negative correlations in most environments, and 16 paired-trait QTL clusters were identified for these traits, 13 of which (81.2%) showed opposite QTL additive effect directions (positive and negative). Seventeen paired-trait QTL clusters were identified for BW with FS, FM, SI, and LP, nine of which (53%) showed the same QTL additive effect direction, and eight of which (47%) showed opposite additive effect directions.

Overall, the QTL clusters for cotton fibre quality and yield throughout the genome showed that paired traits with significant medium or high positive correlations always had the same QTL additive effect direction, indicating that it could be easy to simultaneously improve these traits. However, the paired traits with significant medium or high negative correlations always had opposite QTL additive effect directions, indicating that it could be difficult to simultaneously improve these traits. The paired traits with no or weak positive or negative correlations had either the same or opposite QTL additive effect directions, indicating that the positive and negative additive effects of QTLs in different clusters might have a mutually neutralising effect with respect to these traits.

These conclusions indicated that genes and QTLs in clusters might be closely linked (Cai *et al*., 2002; Durand *et al*., 2012;Reddy *et al*., 2013; Gu *et al*., 2015; Ku *et al*., 2015; Li *et al*., 2015b;Vikram *et al*., 2015; Zhao *et al*., 2016) or have pleiotropic effects (Li *et al*.; 2016c; Zhao *et al*., 2018; Xie *et a*l., 2018; Yuan *et al*., 2018; You *et al*., 2019) in the same genomic region, which provides an explanation for the significant phenotypic correlations between related traits and linkage drag. However, as in previous reports, especially inbreeding populations, the negative genetic correlations between traits could be broken to a certain degree(Cai *et al*., 2002; Durand *et al*., 2012;Reddy *et al*., 2013; Gu *et al*., 2015; Ku *et al*., 2015; Li *et al*., 2015b;Vikram *et al*., 2015; Zhao *et al*., 2016). This result suggests that genetic linkage maybe the main genetic basis of trait correlation. To break the negative correlations between fibre quality traits and yield traits, the six QTL clusters (*qClu-chr16-1*, *qClu-chr20-1*, *qClu-chr6-2*, *qClu-chr25-2*, *qClu-chr13-2*, and *qClu-chr21-3)* should be considered for fine mapping of QTLs for FS, FL, and LP. Techniques such as further separation of the confidence intervals of QTL clusters via secondary separation populations (Cao *et al*., 2015; Liu *et al.*,2016; Fang *et al*., 2017; Xu *et al*., 2017), mutation with ethyl-methanesulfonate (EMS) and radiation to change genes related to traits (Patel *et al*., 2014; Naoumkina *et al*., 2017), and the use of clustered regularly interspaced short palindromic repeats (CRISPR)technology to edit genes related to traits (Gao *et al*., 2017; Janga *et al*., 2017; Wang *et al.*, 2017b and 2018) could be used to improve fibre quality and yield.

Experimental procedures

Development and phenotypic evaluation of the mapping population

An RIL population derived from two upland cotton cultivars, 0-153 and sGK9708, was evaluated and used throughout this study. Briefly, the cross was made in 2001, and a segregation population consisting of196 $F_{6:8}$ RILs was developed as detailed by Sun *et al*. (2012). From 2007 to 2015, multi environmental evaluations were conducted in 22 different ecological locations across two principal cotton planting areas, Xinjiang Uygur Autonomous Region (XUAR) and Yellow River Valley (YRV). In the years 2007–2010 and 2013, this experiment was described in Zhang's report (Zhang *et al*., 2015b). In 2011, the population was planted in Anyang and Zhengzhou; in 2012, it was planted in Zhengzhou; in 2014 and 2015, it was planted in Anyang, Alaer, Shihezi, and Kuerle. The phenotypic evaluations of the population at each location were conducted in a completely randomised block design of one-row plots with two replicates in the YRV locations. Each row was 5 m long and 0.8 m apart, with 20 plants planted in each row except for2013Anyang, which had only one replicate. In the XUAR locations, two-narrow-row plots were applied; each row was 3 m long and planted in wide/narrow alternating row spacing of 0.66/0.10 m, with 25 plants in each row. Field management was performed according to the local farming practices.

Three fibre quality traits, FL, FS, and FM, and three yield traits, BW, LP, and SI, were evaluated for their phenotypic performances throughout the population. Thirty normally opened bolls were sampled from each plot during the harvesting season in September of each testing year to evaluate the phenotypic performances of the aforementioned traits. Briefly, after the seed cotton samples were weighed and ginned, BW, LP, and SI were evaluated. The methods of testing the fibre quality traits are described in Sun's and Zhang's reports (Sun *et al*., 2012; Zhang *et al*., 2016). One-way ANOVA was used to test the significance of the differences in the traits between the two parents. The descriptive statistics were calculated with SPSS 20.0 software (SPSS, Chicago, IL, USA), and the variance and heritability were analysed with Ici Mapping software (Li *et al*., 2007 and 2015c; Meng *et al*., 2015).

Correlation analysis of the six phenotype traits

The correlations between the different traits in the same environments and between the same traits in different en-

vironments were analysed with SPSS 20.0 software if the *P* value was more than 0.05, the two traits had no correlation. When the *P* value was less than 0.05, if the correlation coefficient was from 0 to 0.3or from -0.3 to 0, the two traits had a weak positive or negative correlation; if the correlation coefficient was from 0.3 to 0.5or from -0.5 to -0.3, the two traits had a medium positive or negative correlation; if the correlation coefficient was from 0.5 to 1or from -1 to -0.5, the two traits had a high positive or negative correlation.

Marker filtration

The SSR markers (Sun *et al*., 2012; Zhang *et al*., 2015c; Jamshed *et al*., 2017), the markers developed by SLAF-Seq (Zhang *et al.*, 2016), and markers discovered using the Intl Cotton SNP Consortium_70k chip (Zhang *et al.*, 2017b) were used to generate the consensus genetic map. Before the construction of the consensus genetic map, the markers were filtered. Markers with no position or more than one position on the genome of upland cotton were removed (Li *et al*., 2015a; Zhang *et al*., 2015b), as were markers with unclear genotyping, those with no polymorphism between the parents, those that were heterozygous in at least one parent, those with a missing rate of more than 40%, and those that showed segregation distortion with a *P* value of the Chi-square test less than 0.001.

Consensus genetic map construction

All the markers remaining after filtration were initially partitioned into linkage groups and located on the chromosomes following the procedures detailed in Zhang's research (Zhang *et al*., 2015a) and according to the reference genome database developed from TM-1 (Li *et al.*, 2015a; Zhang *et al.*, 2015b). High Map software was applied to order the alleles and correct genotyping errors within all chromosomes (w=3, l=200, r=7) (Jansen *et al*., 2001; Liu *et al*., 2014; van Ooijen *et al*., 2011). The method of processing and constructing the consensus map was the same as described in Zhang's research (Zhang *et al*., 2016)

The density of markers in the physical map and genetic map was calculated with Perl script with windows of 500Kb and 5cM, and the picture was drawn by CIRCOS 0.66 (Krzywinski *et al.*, 2009). The segregation distortion of each marker was analysed by the Chi-square test using JoinMap 4.0software (van Ooijen., 2006). The SDM loci showing significance with *P* values between 0.001 and 0.05 were maintained to construct the consensus genetic map, any region on the map with more than three SDM loci was defined as a segregation distortion region (SDR) (Zhang *et al.*, 2013).

QTL and QTL cluster identification for fibre quality and yield

Windows QTL Cartographer 2.5 (Wang *et al*., 2001) was applied to map the QTLs using the composite interval mapping method. QTLs identified consistently in at least three environments were considered to be stable (Sun *et al*., 2012). A positive additive effect indicates that the favourable allele comes from the parent 0-153, while a negative effect indicates that the favourable allele comes from the parent sGK9708. The parameter of the software and the rules of QTL naming were described in Sun's and Zhang's reports (Sun *et al*., 2012; Zhang *et al*., 2016).

The stable QTLs identified were compared with the Cotton QTL db database (http://www.cottonqtldb.org) to determine whether they were novel or had been published in previous studies (Rong *et al*., 2007; Lacape *et al*., 2010; Said *et al*., 2013, 2015a and 2015b; Fang *et al*., 2014)or previous GWAS research (Fang *et al.*, 2017; Huang *et al*., 2017; Sun *et al*., 2017; Wang *et al*., 2017a; Ma *et al.*, 2018). The stable QTLs in the database that shared the same or overlapping confidence intervals as the QTLs found in this study were considered to have been identified in previous studies. Stable QTLs for different traits that shared the same or overlapping confidence intervals were considered to reside in QTL clusters, and the overlapping regions were considered to be the confidence intervals of the QTL clusters.

References

Cai, C.P., Niu, E.L., Du, H., Zhao, L., Feng, Y. and Guo. W.Z. (2014) Genome-wide analysis of the WRKY transcription factor gene family in *Gossypium raimondii* and the expression of orthologs in cultivated tetraploid cotton. *The Crop Journal.*2, 87-101Du, X.M., Huang, G., He, S.P., Yang, Z.E., Sun, G.F., Ma, X.F., et al. (2018) Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits. *Nature Genetics*. 50, 796–802

Fang, D.D., Jenkins, J.N., Deng, D.D., McCarty, J.C., Li, P. and Wu, J.X. (2014) Quantitative trait loci analysis of fibre quality traits using a random-mated recombinant inbred population in upland cotton (*Gossypium hirsutum* L.). *BMC Genomics*. 15, 397

Jamshed, M., Jia, F., Gong, J.W., Palanga, K.K., Shi, Y.Z., Li, J.W., Shang, H.H., *et al*. (2016) Identification of stable quantitative trait loci (QTLs) for fibre quality traits across multiple environments in *Gossypium hirsutum* recombinant inbred line population. *BMC Genomics*. 17, 197

Jansen, J., de Jong, A.G. And van Ooijen, J.W. (2001) Constructing dense genetic linkage maps. *Theor Appl Genet*. 102, 1113-22

Li, F.G., Fan, G.Y., Wang, K.B., Sun, F.M., Yuan, Y.L., Song, G.L., Lin, Q. *et al*. (2014) Genome sequence of the cultivated cotton *Gossypium arboretum*. *Nature Genetics*. 46, 567-72.

Li, F.G., Fan, G.Y., Lu, C.R., Xiao, G.H., Zou, C.S., Kohel, R.J., Ma, Z.Y. *et al*. (2015a) Genome sequence of cultivated upland cotton (*Gossypium hirsutum* TM-1). *Nature Biotechnology*. 33, 524-30. Paterson, A.H., Wendel, J.F., Gundlach, W.H., Guo, H., Jenkins, J., Jin, D., Llewellyn, D. et al. (2012) Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature*. 492, 423-7Rong, J.K., Abbey, C., Browers, J.E., Brubaker, C.L., Cheng, C., Chee, P.W., Delmonte. T.A. *et al*. (2004) A 3347-Locus Genetic Recombination Map of Sequence-Tagged

Sites Reveals Features of Genome Organization, Transmission and Evolution of Cotton (*Gossypium*). *Genetics*. 166, 389-417

Sun, F.D., Zhang, J.H., Wang, S.F., Gong, W.K., Shi, Y.Z., Liu, A.Y., Li, J.W. *et al*. (2012) QTL mapping for fibre quality traits across multiple generations and environments in upland cotton. *Mol Breeding*. 30, 569-82.

Wang, M.J., Tu, L.L., Yuan, D.J., Zhu, D., Shen, C., Li, J.Y., Liu, F.Y. et al. (2019) Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*. *Nature Genetics.* 51, 224–229

Zhang, T.Z., Hu, Y., Jiang, W.K., Fang, L., Guan, X.Y., Chen, J.D., Zhang, J.B. *et al*. (2015b) Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fibre improvement. *Nature Biotechnology*. 33, 531-7

Zhang, Z., Li, J.W., Jamshed, M., Cai, J., Jia, F., Shi, Y.Z., Gong, J.W. *et al*. (2015c) High resolution consensus mapping of quantitative trait loci for fber strength, length and micronaire on chromosome 25 of the upland cotton (*Gossypium hirsutum* L.). *PLoS ONE*. 10, e0135430.

Zhang, Z., Shang, H.H., Shi, Y.Z., Huang, L., Li, J.W., Ge, Q., Gong, J.W. *et al*. (2016b) Construction of a high-density genetic map by specific locus amplified fragment sequencing (SLAF-seq) and its application to quantitative trait loci (QTL) analysis for boll weight in upland cotton (*Gossypium hirsutum*.). *BMC Plant Biology*. 16, 79

Zhang, Z., Ge, Q., Liu, A.Y., Li, J.W., Gong J.W., Shang, H.H., Shi, Y.Z., *et al*. (2017b) Construction of a High-density Genetic Map and Its Application to QTL Identification for Fibre Strength in Upland Cotton. *Crop Science*. **57**, 774-788

ICAC Combines Three Reports into One Massive, Annual 'Cotton Data Book'

The new ICAC Cotton Data Book:

- Is an annual compilation of three reports that had previously been published once every three years
- Will be published in December of each year
- Sells for $250 less than the three reports combined sold separately in prior years

Visit www.icac.org to purchase your copy!