



HAL
open science

Cotton genetic resources. A review

Mehboob-Ur-Rahman, Shaheen, Tabbasam, Muhammad Iqbal, Ashraf, Zafar,
Andrew Paterson

► **To cite this version:**

Mehboob-Ur-Rahman, Shaheen, Tabbasam, Muhammad Iqbal, Ashraf, et al.. Cotton genetic resources. A review. *Agronomy for Sustainable Development*, 2012, 32 (2), pp.419-432. 10.1007/s13593-011-0051-z . hal-00930526

HAL Id: hal-00930526

<https://hal.science/hal-00930526>

Submitted on 11 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Cotton genetic resources. A review

Mehboob-ur-Rahman · Tayyaba Shaheen · Nabila Tabbasam ·
Muhammad Atif Iqbal · Muhammad Ashraf · Yusuf Zafar · Andrew H. Paterson

Accepted: 29 July 2011 / Published online: 13 October 2011
© INRA and Springer-Verlag, France 2011

Abstract Since 6000 BC, cotton has been cultivated for lint fiber, which now dominates the natural textile industry worldwide. Common resources such as an integrated web database, a microsatellite database, and comparative quantitative trait loci (QTL) resources for *Gossypium* have accelerated the progress towards quantifying the impact of repeated human dispersals and selection regimes on various gene pools of the genus *Gossypium*. Out of 50 *Gossypium* species, four have been domesticated—two diploids and two tetraploids—for elimination of hard seed coat, improvement in lint percentage of about 40% and fiber length of 22%, larger boll size, and day-neutral reproductive habit. The major drawback of domestication is the lack of genetic diversity. This lack of genetic diversity is observed more in *Gossypium hirsutum* L. cultivars characterizing upland cotton than in *Gossypium barbadense*, typical of Pima and Egyptian cotton.

Much of the genetic diversity among *G. barbadense* cultivars is attributed to the introgression of *G. hirsutum* alleles. This process highlights the importance of introgression of new alleles from accessions of all the *Gossypium* species into cultivated cotton species. Among the genomic resources, about 16,162 publicly available SSRs and 312 mapped cotton RFLP sequences containing simple sequence repeat (SSR), restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), and random amplified polymorphic DNA (RAPD) markers have been surveyed on numerous mapping populations, and developed about 26 linkage maps (SSR, RFLP, AFLP, and RAPD). Reports show the identification of DNA markers associated with over 29 important traits or QTLs such as fiber quality and yield, leaf and flower morphology, trichome density and their distribution, and disease resistance. In comparative mapping studies, 432 QTLs mapped on 11 different mapping populations were aligned on a high-density reference map containing 3,475 loci. In a meta-analysis study of over 1,000 QTLs obtained from backcross population and recombinant inbred line populations derived from the same parents, most consistent meta-clusters were reported for fiber color, fineness, and length. For exploring the function of genes, the targeting induced local lesions in genomes (TILLING) approach—avoiding gene transfer process was used for identifying a brassinosteroid receptor gene that is involved in fiber development. Lastly, cotton genome has been enriched with genes isolated from distantly related organisms using various transformation methods. For example, *CryIac*, *CryIAb*, and herbicide-resistant genes were transformed in cotton that covered a vast majority of cotton acreage worldwide. Here the authors discuss investigations for improving the efficacy of transformation and regeneration systems, and for searching new genes or silencing the unwanted cotton genes using RNAi technology. We suggest

Mehboob-ur-Rahman (✉) · T. Shaheen · N. Tabbasam ·
M. A. Iqbal · Y. Zafar
Plant Genomics and Molecular Breeding Lab, National Institute
for Biotechnology and Genetic Engineering (NIBGE),
P.O. Box 577, Jhang Road,
Faisalabad, Pakistan
e-mail: mehboob_pbd@yahoo.com

M. Ashraf
Department of Botany, University of Agriculture,
Faisalabad, Pakistan

A. H. Paterson
Plant Genome Mapping Laboratory, University of Georgia,
Athens, GA 30602, USA

T. Shaheen
Department of Bioinformatics and Biotechnology,
Government College University,
Faisalabad, Pakistan

initiating projects on sequencing the diploid and tetraploid genomes for exploring the extent of genetic variations, developing TILLING populations, initiating nested association mapping studies, and developing third generation genetically modified cotton, collectively setting the stage for sustaining cotton production under continually changing production conditions, climates, and human needs.

Keywords *Gossypium* · Genetic diversity · Diploid cotton · Tetraploid cotton · Hybrid cotton · Structural genomic resources · Functional genomic resources · Dry lab resources · Translational genomics · Comparative mapping · DNA marker · RFLP · RAPD · SSRs · SNPs · QTLs · Marker-assisted selection · GM cotton

Abbreviations

AFLP	Amplified fragment length polymorphism
BC	Backcross population
CLCuD	Cotton leaf curl disease
GUS gene	Beta-glucuronidase gene
LD	Linkage disequilibrium
MAS	Marker-assisted selection
MYA	Million years ago
NAM	Nested association mapping
PCR	Polymerase chain reaction
QTLs	Quantitative trait loci
RAPD	Random amplified polymorphic DNA
RFLP	Restriction fragment length polymorphism
RIL	Recombinant inbred line
SNPs	Single nucleotide polymorphisms
SSR	Simple sequence repeat
TILLING	Targeting induced local lesions in genomes

Contents

1. Introduction	2
2. Spectrum of genetic diversity	3
2.1 Genetic diversity at inter-specific level	3
2.2 Genetic diversity at intra-specific level	3
3. Genomic Resources	4
3.1 Conventional structural genomic resources	4
3.1.1 Cytogenetic stocks	4
3.1.2 Genetic maps	5
3.1.3 Scope of developing new populations	5
3.1.3.1 Utilizing untapped resources	5
3.1.3.2 New mapping approaches – nested association mapping	5
3.2 Functional Genomic Resources	7
3.2.1 Knock-out populations	7
3.2.2 Opportunities	7

4. Dry-lab resources—comparative genomics	7
4.1 Transfer of genomic information from model species to cotton	7
4.2 Broader Applications	9
4.2.1 Identification of genes for cotton breeding	9
4.2.2 Utilization of $\alpha\omega$ genetic resources	9
5. Implications	9
5.1 Knowledge-based breeding—proactive decision	9
5.2 Marker-assisted selection (MAS) for various traits of interest	10
5.3 Fiber genes—exploring unique genetic pathways	10
5.4 Translational genomics	10
5.5 Genetic resource for other species	10
6. Conclusion	11
7. References	11

1 Introduction

Cotton is a world leading natural fiber crop sustaining one of the world's largest industries (textiles) with an annual worldwide economic impact of about \$500 billion. Genetic resources of cotton encompassing a range of germplasm of ~50 cotton species offers unique opportunities for understanding the evolution of cultivated cotton from wild ancestors which will pave the way for dissecting the evolution of lint fiber and also the role of polyploidy for enhancing the lint production. Other features like understanding the extent of genetic diversity and its utilization in improving sustainability of cotton production and lint quality and also the utilization of bio-based alternatives to petrochemicals are being explored at length.

In this article, advances made during the last two decades have been summarized. For example, progress in understanding variations in multiple morphological, physiological, biochemical, and genetically important traits have been discussed. Like other important crop species, much emphasis was given on understanding the cotton genome of the genus *Gossypium* using various cutting edge genomic technologies which substantially laid down the foundation for sustaining lint production worldwide (Rahman et al. 2009). Still there are many aspects to explore (Chen et al. 2009). Tremendous domestication has resulted in narrow genetic base of cotton (Rahman et al. 2008). Conventional genetic resources have been utilized over the years; however, a decline in cotton production has been seen for last few

decades (Helms 2000). It was the need of time to generate new resources to combat the problems of new era. Genomic tools have been deployed extensively for improving fiber features, developing cotton cultivars conferring resistance to insect pests and diseases, and abiotic stresses. It has been advocated that genetic manipulation was a revolution in improvement of genomes (Abelson 1998). This review covers all available conventional and modern genetic resources in cotton, and their potential for utilization in the future. Improved utilization of the available genetic resources can address the encountered threats to cotton production today.

2 Spectrum of genetic diversity

Cotton (*Gossypium* spp.; Malvaceae family) contributes to the world economy as a leading fiber crop and significant oilseed. The genus *Gossypium* comprises of about 50 species (45 diploids $2n=2\times=26$; and 5 tetraploids $2n=4\times=52$, Fryxell et al. 1992). Tetraploid species evolved ~1–2 million years ago (MYA) after the hybridization of the A and “D” genome species, diverged from a common ancestor about 4–11 MYA (Wendel 1989). Each of the two subgenomes represent one copy of almost all genes thus tetraploid cotton harbor two or more copies of each gene (Rong et al. 2004). Also the order of these genes remained like their diploid ancestors (Brubaker et al. 1999). The tetraploid species *Gossypium hirsutum* L. and *Gossypium barbadense* L. (denoted hereafter as *GH* and *GB*, respectively), represent the most cultivated cotton species. Crosses can be made between the cultivated cotton species and their F1s can further be crossed with wild tetraploid species (*Gossypium darwinii*, *Gossypium tomentosum*, and *Gossypium mustelinum*, denoted hereafter as *GD*, *GT*, and *GM*, respectively) which result in normal hybrids and they can further produce several fertile progenies (Waghmare et al. 2005).

2.1 Genetic diversity at interspecific level

Present day cotton is the outcome of rigorous selection for production of desired type of fibers for the improved harvesting and processing. The side effect of this selection was the narrow genetic diversity for some specific traits including resistance against drought (Rosenow et al. 1983).

One solution of this problem is to bring alleles from wild ancestors under cultivation which can provide fundamental genetic solutions to various agricultural trials (Gur and Zamir 2004). Exotic tetraploid cottons are relatively more tolerant to drought and heat. For example, *GT* is native to

Hawaii, and is found in rocky, arid, or clay coastal plains (http://kalama.doe.hawaii.edu/hern95/pt009/Ann/mcc_nativeplants.html). Progenies generated after making an interspecific cross between *GT* and *GH* showed relatively high tolerance under water stress (Gotemare and Singh 2004). *G. darwinii*, endemic to the Galapagos Islands, survived naturally in the arid islands (Fryxell 1979). *G. mustelinum* is widespread in tremendously dry areas (Pickersgill et al. 1975) of few states located in the semi-arid region of Northeastern Brazil (Silva et al. 2006). Due to better adaptation to drought stress, an exotic strain of *GH*, var. *marie-galante* remained under cultivation in the beginning of twentieth century in Brazil (Boulanger and Pinheiro 1971). In another study, substantial differences in heat tolerance, root growth, WUE, and dry matter accumulation among exotic *GH* lines were shown (Quisenberry et al. 1982). However, different degrees of sterility and distinct segregation distortions, and propensity to preserve the parental haplotypes of interspecific hybrids are unusual features which hinder the progress of enriching the genomes of cultivated cotton species (Jiang et al. 2000b; Waghmare et al. 2005).

2.2 Genetic diversity at intraspecific level

Presently, cotton yield improvement trends show the effect of a variety of biotic and abiotic threats on cotton due to meager genetic diversity. This inertia in the growth of cotton production, primarily of *GH*, was the compelling factor for making a “Blue Ribbon Committee” of public and private sector cotton scientists for finding reasons of depression in yield, and also to design a strategy to combat the problem (Helms 2000). This committee utilized the yield data of largely two public trials (National Cotton Variety Tests, Rayburn et al. 1999, and National Agricultural Statistics Service, USDA). A linear model shows that during the past 39 years cotton production have raised ~6.7 kg ha⁻¹ year⁻¹ (1.3% annual growth rate) while a polynomial model elucidates that this increment has not been linear. Cotton yield increased from 1970 through 1985. However, the rate of change in cotton yield deteriorated from 1992 through 1998 in the period of 1985 through 1998, thus the actual yields declined. At the start of twenty-first century, world average cotton yield has been raised from 566 kg/ha in 1998/1999 to a record of 793 kg/ha in 2007/2008. Numerous factors were involved in this marvelous increment like new technologies, effective utilization of available resources, and shifts in cotton production areas. However, since 2007/2008 the cotton yield has deteriorated. The world cotton yield has been projected down by 5% to 725 kg/ha in 2009/2010, the lowest in six seasons. This decline in cotton production over the last three seasons has been caused by the meager production in USA and China. The cotton production in

USA fell by 43% to 2.7 million tons, while Chinese production depressed by 15% to 6.8 million tons. (http://www.abare.gov.au/outlook/_download/fibres_gruere.pdf). This deterioration in yield and year-to-year differences in production, were ~4 times higher in the era of 1980–1998 than in 1960–1979. High instability in yield illustrates a threat for the grower, due to year-to-year uncertainty in the size and quality of the crop.

There is a meager genetic variability in the cotton gene pool employed in cultivar development due to repeated utilization of a few genetic backgrounds for new cultivars development (May et al. 1995; Rahman et al. 2002; 2005; 2008). Moreover, farmers have planted the similar germplasm on large areas which lead to high level of genetic homogeneity in field. For example, by 1995, field uniformity had reached as high as 38% in the mid-south US production area (Van Esbroeck et al. 1998). This already high level of genetic uniformity has been aggravated by the prevalent (ca. 76% of 2003 US hectareage) cultivation of cultivars which are transgenic and resulted from a backcross (BC) breeding with a few closely related genetic backgrounds, causing a high degree of genetic uniformity on producer farms.

The use of exotic accessions for broadening the genetic base of cultivated cotton has barely been explored—although many accessions with a host of valuable attributes have been converted to “breeder-friendly” forms. Exotic cottons have been shunned by mainstream breeding programs, because most are photoperiod-sensitive and flower too late in the season for temperate production (McCarty and Jenkins 1992). In fact, no exotic germplasm appears in the pedigrees of modern cotton cultivars (Calhoun et al. 1997). Early maturity also avoids hazards such as insects, diseases, drought, and cold (cf. Adkisson et al. 1982). Flowering of cotton is under multigenic control, but with high heritability (0.62 for maturity date: Quisenberry 1977), and partial dominance of day-neutrality (Kohel et al. 1974).

The difficulties in using photoperiodic cottons for breeding impelled a “Cotton Conversion Program” (McCarty and Jenkins 1992), which has now “converted” (backcrossed) numerous exotic cottons. Genotypes were prioritized for conversion based on specific valuable attributes, and also provide a good sampling of the geographic distribution of diversity in *G. hirsutum*. Exotic cottons are crossed to a day-length insensitive flowering donor parent (mostly DPL16), then backcrossed to the exotic parent followed by self-pollination, for a total of four cycles (eight generations). Backcrosses are made in the winter at a tropical site where the exotic parent can flower. Selection for day-neutral (temperate) flowering is practiced among BC progeny at a temperate site. In total, “conversion” takes about 5 years, and is predicted to reconstitute

ca. 96% of the exotic (recurrent) genome, except in genomic regions that confer short-day flowering.

Presently, no conversion lines have been developed from the pedigrees of any recent cultivars or improved germplasm. Since these lines have been available since 1979, this is not simply a result of a slow breeding cycle. It has been due to a lack of exploration of the lines by mainstream breeders, due largely to reluctance to deal with the complexities associated with exotic germplasm manipulation.

An additional problem in designing effective strategies to sample these exotics for further utilization is that their relationships to one another are poorly understood. A selection of converted race stocks was found to be more related to the *G. hirsutum* genetic standard (TM-1) than predicted, after four backcrosses to the corresponding primitive accession; however, these conclusions are based on only 56 simple sequence repeats (SSRs) distributed across a large (>4,000 cM), complex genome (Liu et al. 2000b). Other efforts to categorize the exotic germplasm have grouped them into botanical races based on morphological and geographic criteria (Fryxell 1979); however, these races, and especially some groupings that are argued to represent species such as “*Gossypium lanceolatum*,” are not well-supported by the very limited neutral (molecular) marker data that exists (Brubaker and Wendel 1993).

It is also noteworthy that exotic chromatin is lost from advanced-backcross progenies more quickly than would be expected (Jiang et al. 2000b; Liu et al. 2000b). This suggests that the few classical studies of cotton introgression that have been done have under-estimated the potential benefits of introgression due to the rapid loss of introgressed chromatin. This potential artifact is added to the problem of confounding the benefits of introgression with the penalties, due to past inability to evaluate the effects of individual chromosome segments (lacking DNA markers). The use of a detailed molecular map will resolve these difficulties, identifying rare individuals retaining the exotic chromatin, thus permitting the independent evaluation of the positive and negative effects of each introgressed chromosome segment.

3 Genomic resources

3.1 Conventional structural genomic resources

3.1.1 Cytogenetic stocks

Aneuploid substitution stocks have been developed from tetraploid TM-1 (*GH*) × 3–79 (*GB*) (Endrizzi and Ramsay 1979) and TM-1 × *GT* (Saha et al. 2006). Three of the chromosomes remained unidentified while 23 of the 26

chromosomes were identified by monosomics and telosomics together. Numerous SSRs and restriction fragment length polymorphism (RFLPs) have been assigned to respective chromosomes utilizing these resources.

3.1.2 Genetic maps

There exists at many published genetic maps to map useful traits and markers which includes ~5,000 markers in public database including 3,300 RFLP, 700 amplified fragment length polymorphism (AFLP), 1,000 SSR, and 100 single nucleotide polymorphism (SNP). The STS-based genetic maps of diploid (D) and tetraploid (AtDt¹) can serve as a source of probes for many of the other maps. Here “At” stands for A subgenome while “Dt” stands for D subgenome in tetraploid cotton *Gossypium* genomes. The reference maps include 2,584 loci at 1.72 cM (~600 kb) intervals comprised on 2,007 probes (AtDt); and 1,014 loci at 1.42 cM (~600 kb) intervals identified by 809 probes (D) (Rong et al. 2004; Rong et al. 2005a). Respective genomes show great collinearity with each other (Rong et al. 2004) to get benefit of this, the gene order was concluded of a proposed mutual ancestor of the At, Dt, and D genomes. In total, 3,016 loci recognized by 2,337 probes were included in this map, and it spanned 2,324.7 cM. With the help of three different populations derived from interspecific crosses supplementary marker rich maps were developed which have been extensively utilized as reference maps for studying quantitative trait loci (QTLs) (Lacape et al. 2003; Guo et al. 2007; Lacape et al. 2007; Yu et al. 2007).

Diagnostic DNA markers linked with fiber traits have also been identified with the help of cotton molecular maps (see Rahman et al. 2009; 2011; Cai et al. 2010 for a review). Study of genomics of drought tolerance has become vital due to heavy water use of cotton (Saranga et al. 2001; Saranga et al. 2004; Zhao et al. 2008; Ullah 2009). During the last few years, major focus has also been the disease resistance with published information for *Xanthomonas* (Wright et al. 1998; Rungis et al. 2002), *Verticillium* (Bolek et al. 2005; Wang et al. 2008; Yang et al. 2008), root knot nematode (Shen et al. 2006; Wang et al. 2006; Wang and Roberts 2006a, b; Ynturi et al. 2006; Niu et al. 2007), *Thielaviopsis* (Niu et al. 2008), and cotton leaf curl disease (CLCuD; Fig. 1, Rahman et al. 2005, 2006).

Nuclear restorer of a cytoplasmic male sterility (Fig. 2) has been a focus of R&D efforts because of increasing interest in growing hybrid cotton (Guo et al. 1998; Lan et al. 1999; Zhang and Stewart 2004; Wang et al. 2007b). Different traits of defense umbrella such as leaf morphology (Jiang et al. 2000a; Song et al. 2005; Waghmare et al. 2005; Mei et al. 2004; Hao et al. 2008) and leaf color (Fig. 3, Ali et al. 2009b) along with the pubescence, a specific characteristic of *GH* (Wright et al. 1999; Lacape and Nguyen 2005;

Desai et al. 2008; Ali et al. 2009a), have also been the focus of study. Physical characteristics of seed and its nutritional value is also being mapped due to worth of cotton seed (Song and Zhang 2007). There is possibility that this list has some discrepancy (and the authors apologize for any unintentional oversights).

A more profound method for the study of genetic control of a trait is the alignment and evaluation of several QTL mapping experiments using a mutual reference map rather than single study; it elucidates the genomic arrangement of trait distinction. A dense genetic map containing 432 QTLs (for fiber quality and its yield, leaf and flower morphology, trichome density, and their distribution etc.) and 3,475 loci were identified in a total of 11 populations (Jiang et al. 1998, 2000a; Wright et al. 1999; Saranga et al. 2001, 2004; Paterson et al. 2003; Chee et al. 2005b; Chee et al. 2005a; Draye et al. 2005; Rong et al. 2005b; Waghmare et al. 2005). Later, these QTLs were mapped on a consensus map which was presumed to be comparable to the DNA marker organization of the hypothetical ancestors of the two “subgenomes” of the tetraploids (Rong et al. 2005b). In another meta-analysis study of over 1,000 QTLs obtained from BC and recombinant inbred line (RIL) populations derived from the same parents, most consistent meta-clusters were reported for fiber color, fineness, and length (Lacape et al. 2010).

Assumption of cotton–*Arabidopsis* synteny relationships has become convenient due to consensus maps and it assists the study of correspondence between fiber or trichome-related *Arabidopsis* genes and the cotton QTLs. This CMap resource was generated and can be accessed at <http://chibba.agtec.uga.edu/cgi-bin/cmap/viewer> for using in various genomic studies.

3.1.3 Scope of developing new populations

Utilizing untapped resources Development of genetic maps with molecular markers needs mapping populations. Numerous linkage maps have been reported which are based on intraspecific crosses between upland cottons (Zuo et al. 2000; Ulloa et al. 2002; Shen et al. 2005; Zhang et al. 2005; Wang et al. 2006; Guo et al. 2007; Shen et al. 2007; Wang et al. 2007a; Ma et al. 2008; Ali et al. 2009a, b) but these maps could not provide extensive genome coverage because of the narrow genetic base resulted due to domestication of *GH*. However, the hybrids derived from crossing *GH* and *GB* have shown a promising scenario because of finding large number of polymorphisms. Mostly available larger genetic maps of cotton are developed from these interspecific populations (see Rahman et al. 2009; 2011 for a review).

New mapping approaches—nested association mapping Genetic mapping is based on linkage disequilibrium (LD).

Fig. 1 **A** Healthy cotton plant, **B** healthy cotton leaf, **C** infected cotton plant. With cup shaped leaf structure, **D** infected leaf (1 enation, 2 large veins thickening, 3 small veins thickening)



A: Healthy cotton plant



C: Infected cotton plant
With cup shaped leaf structure



B: Healthy cotton leaf



D: Infected leaf (1= Enation; 2= Large Veins thickening; 3= Small veins thickening)

Artificial populations, usually from crosses between two inbred parents, are in very strong linkage disequilibrium. Natural populations are usually the products of many cycles of recombination, and have the potential to pinpoint QTLs to high resolution using “association mapping” (Varshney et al. 2006). Mapping of gene which control complex traits is done with nested association mapping (NAM) which is comparatively a new approach (Fig. 4). In this method the statistical power of mapping QTLs and the high (potentially gene-level) chromosomal resolution of association mapping is combined. Populations used in NAM are “nested” as these populations share one parent common. However, each population has another alternate parent. In NAM, a possible number of (e.g., about 1,000–2,000) mutual SNPs which are parent specific are genotyped along with genotyping of the parents with a super resolution (e.g., by sequencing the entire genome). Common parent-specific SNPs are used to characterize each chromosomal segment defined in each segregant according to whether it derives from the common parent or the alternate parent. Like this, high-resolution

sequence data can be concluded from the parents onto the segregating progeny. In this way an association analysis can be performed on the whole population then. This method sometimes permits one to map the genes controlling a specific trait, to single-gene-level resolution. Yu et al. (2008) suggested a nested NAM strategy which utilizes information from genome sequence of RIL populations which are developed after numerous crosses of parental inbreds. This QTL mapping approach using information from linkage and LD is supposed to be effective in detecting QTLs in genome-wide approaches and a better mapping resolution when linkage and LD information are utilized simultaneously (Yu et al. 2008).

Association mapping which is based on LD is an influential molecular tool for dissection and exploitation of the natural genetic diversity which can further be utilized in marker-assisted selection (MAS). Availability of beneficial genetic variability among the exotic germplasm of *G. hirsutum* has been verified with the genome-wide LD and association mapping for lint quality traits with SSRs. These

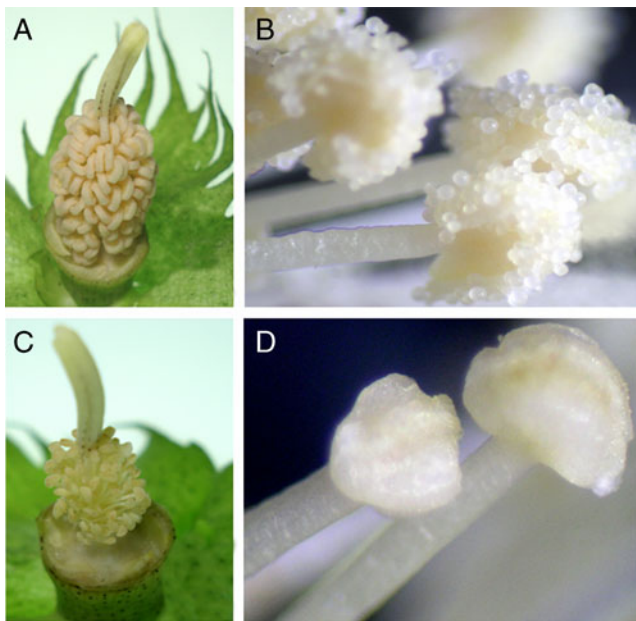


Fig. 2 Comparison between fertile and sterile flowers are shown **A** fertile anthers, **B** fertile pollen grains, **C** sterile anthers, **D** pollen grains are not present on the surface of pollen sacs

outcomes validate the utility of association mapping for efficiently utilization of new genetic variation in MAS programs (Abdurakhmonov et al. 2008).

3.2 Functional genomic resources

3.2.1 Knock-out populations

Targeting induced local lesions in genomes (TILLING) is an effective and widely applicable approach for identifying the functions of genes through developing a knock-out population by exposing the seed with a mutagen such as EMS. Preliminary work has been done to introduce TILLING in diploid cotton (*Gossypium arboreum*) (Auld et al. 2009).



Fig. 3 Comparison between green and red leaf color. **a** Green leaf color, **b** red leaf color

The TILLING approach was used in identifying a brassino steroid receptor gene which is very important in fiber development. The unpublished results indicated the high applicability of TILLING in cotton.

Insertional mutagenesis by transposon or T-DNA-based gene tagging is another effective tool for reverse genetics. Transposon gene tagging is being widely utilized in many plants (Auld et al. 2009). Cotton genetic stocks containing the Ac/Ds transposon system from maize have been developed, but have not been well-characterized yet (N. Trolinder and R. Wright, personal communication).

3.2.2 Opportunities

Transposon tagging and TILLING populations can be utilized for increasing the allelic diversity and improvement of cotton germplasm by expanding genetic resources which would open new avenues in genetic research. Knock-out populations can be screened for novel phenotypes which affect cotton growth and development (Auld et al. 2009). Genetic diversity is important in context of combating encountered diseases like CLCuD (Rahman et al. 2005). Availability of diverse lines with contrasting traits will increase the chance of developing virus resistant lines.

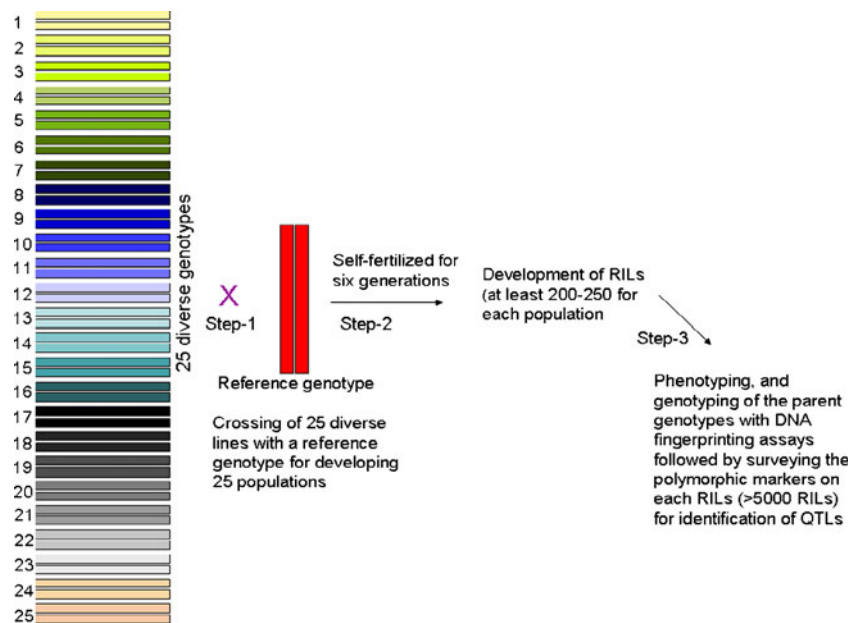
4 Dry lab resources—comparative genomics

4.1 Transfer of genomic information from model species to cotton

It is assumed that *Malvales* (including cultivated cotton species) and *Brassicales* (including *Arabidopsis*) share a common ancestry ~80 MYA (Bowers et al. 2003). The course of evolution occurred through functional modifications, deletion of sequences, and genomic rearrangements. During the homology study of EST obtained from *Gossypium raimondii*, *G. arboreum*, and *G. hirsutum* cDNA libraries it was found that about 55–70% of ESTs were found to have homologs in the *Arabidopsis* genome while rest of the 30–45% of cotton genes were found to be absent or unrecognizable in *Arabidopsis*. Similarity in gene arrangement in cotton and *Arabidopsis* was derived utilizing these homologs (Rong et al. 2005b). Consequently, with the help of two analytical approaches about half of the cotton genome was found to be syntenic with *Arabidopsis*.

Fiber which is the elementary reason of cotton cultivation has been divided into two components, “fuzz” and “lint” (Fryxell 1963) depending on their length and origination. Like seed hairs of wild cotton fuzz fibers are short in length (<5 mm) and attached to the seed coat tightly. Lint fibers are economically important as they are much longer and during ginning can be separated from the

Fig. 4 Schematic illustration of development of nested association mapping (NAM) populations. *RIL* recombinant inbred line, *QTL* quantitative trait loci



seed coat easily. Based on the average length (“staple”), lint fibers are classified as: short staple (<21 mm) in *G. arboreum* and *Gossypium herbaceum*; long staple (28–32 mm) in *G. hirsutum* (“upland”); and extra-long staple (greater than 35 mm) in *G. barbadense*. They usually resemble with *Arabidopsis* trichomes.

Identification of genes associated with the development of fiber and its quality is very complicated because of the large genome size of cotton [2,246 mega bases (Arumuganathan and Earle 1991)] and hindrances in its transformation. Similarities between cotton fibers and *Arabidopsis* trichomes can play a key role in the efforts to clone cotton genes.

There are numerous reports on study of functions of trichome genes and models for trichome development regulation in *Arabidopsis* (Marks 1997; Larkin et al. 2003; Hulskamp 2004; Ramsay and Glover 2005; Jacoby et al. 2008; Marks et al. 2009; Plett et al. 2009). Cotton *MYB* gene is an example of utilizing *Arabidopsis* information in cotton. Many *MYB* genes with R_2/R_3 repeats such as *GL1*, *GL3*, and *WER*, serve as positive regulators (Walker et al. 1999), while others like *TRY* and *CPC* act as negative regulators (Schellmann et al. 2002; Wada et al. 2002). A total of three functional domains including a basic DNA-binding domain (DBD), a negative regulatory domain, and an acidic transacting domain are the basic feature of *MYB* genes. Due to conserved nature of DBD which consist of two or three imperfect repeats (R_1 , R_2 , and R_3), it was utilized for designing polymerase chain reaction (PCR) primers for cotton *MYB* genes amplification. This approach was helpful in identification of 65 cotton *MYB* genes some of them include fiber-related genes (Loguercio et al. 1999; Suo et al. 2003; Hsu et al. 2005). The most important gene

is *GhMYB109* as its expression is specific to fiber initiation and its elongation. This gene has similarity with *AtWER* and *AtGL1* in sequence (Suo et al. 2003). In patterning of trichomes both *GL1* and *GaMYB2* are involved. *GL1* acts as an enhancer in trichomes while it acts as a repressor in nontrichome cells, thus conferring a trichome-specific expression of *MYB* gene. *MYB109*, amplified from *G. arboreum*, showed similarity in function with *GaMYB2* derived from *Arabidopsis* trichomes (Wang et al. 2004).

Other genes like *GhTTG1* and *GhTTG3* which are fiber related have been cloned from cotton with the help of *Arabidopsis* trichome gene sequence. Trichome initiation in *Arabidopsis* is controlled by the gene *TTG1*, containing four WD-40 repeats, forming an initiation complex with *GL1* and *GL3*. Four homologs of *AtTTG1* were isolated and characterized from cotton with the help of a PCR-based method in combination with library screening procedures. These four sequences have been further characterized into two different pairs, one pair, *GhTTG1* and *GhTTG3*, showed 80% similarity in amino acid sequence with the *AtTTG1*, and the second pair, *GhTTG2* and *GhTTG4*, has about 62% similarity. The outcome of the experiment of testing functionality in *Arabidopsis* of these four cotton *TTG* genes was that both *GhTTG1* and *GhTTG3* have the ability to complement *ttg1* mutant (Humphries et al. 2005). This mutant has a number of defects such as lack of trichomes and pigments in seed coat, while the stable transformation of *GhTTG1* and *GhTTG3* restored the normal development of trichomes. These findings are promising for the functionality of such WD-repeat proteins in cotton fiber initiation. Furthermore, expression of *LTP3* and *LTP6*, which are cotton lipid transfer protein genes, has been found to be fiber cells specific (Hsu et al. 1999; Liu et

al. 2000a). Transformation of the constructs of their promoters with beta-glucuronidase genes (GUS genes) into tobacco plants has shown expression of GUS gene only in trichomes, signifying the trichome-specific expression of *LTP3* and *LTP6*. Based on these results, cell growth regulators of fibers can be functionally characterized through utilizing the “similarities at DNA level” of cotton with the other model species like tobacco and *Arabidopsis* (Rong and Paterson 2009).

Another achievement of comparative genomics of cotton and *Arabidopsis* is a comparison of positions of fiber-related QTLs and mutants with cloned fiber candidate genes with the help of cotton–*Arabidopsis* synteny (Rong et al. 2007). Numerous such parallels were detected which are available at CMap database (www.plantgenome.uga.edu/cmap). A significant correlations between the number of *Arabidopsis* candidate genes with the number of QTLs involved in fiber development in cotton were found (Rong et al. 2005b) which may have utility in cloning genes from cotton,

4.2 Broader applications

4.2.1 Identification of genes for cotton breeding

Model plants like *Arabidopsis* are a source of information for many metabolic pathways, i.e., in *Arabidopsis* tolerance to salt (Tester and Davenport 2003), floral development of flowers (Running and Hake 2001), and resistance to diseases (Cunnac et al. 2007). These are vital genes of agricultural importance which can be identified by making comparisons of linkage maps and or other genomic information between the *Arabidopsis* and cotton. Availability of full genome sequence of *Arabidopsis* can be helpful in saturating the map of a desired region in cotton with detection of cotton ESTs homologous to *Arabidopsis* genes in the region.

4.2.2 Utilization of $\alpha\omega$ genetic resources

$\alpha\omega$ genetic resources of cotton are genes or part of genes taken from any source and placed in cotton (Stewart 1994). The *Bt* gene, for example, was isolated from bacterial species *Bacillus thuringiensis* and transferred into cotton. Resistance against herbicide bromoxynil is caused by the nitrilase gene. This gene was obtained from bacterial species belonging to the genus *Klebsiella*. Resistant genes (acetolactate synthase) for imidazolinone and sulfonylurea classes of herbicide were obtained from both the genomes of higher plants and bacteria (Stewart 1994). In Pakistan, biolistic gene transformation of cotton was experimented which was not found very productive (Haq et al. 2005). In another experiment, a silicon carbide whisker-mediated

gene transfer system, established for improving the recovery of fertile and stable transformants of *GH* cv. Coker-312 (Asad et al. 2008), was found much more practical than after the particle bombardment system in cotton.

Bioengineered cotton varieties which are commercially known as Bollgard cotton have the ability of production of an insecticidal protein from *Bacillus thuringiensis*. This protein is effective against some lepidopterous cotton insect pests when produced in cotton. Bollgard cottons (first generation—containing *Cry 1Ac* gene, and second generation, containing *Cry 1Ac* gene and *Cry 2Ab* gene) have been first commercialized in 1996, presently covered ~62% of the cultivated cotton area worldwide. Cotton production has become less costly due to ability of this product to replace chemical insecticides for the control of cotton bollworm, *Helicoverpa zea*, *Heliothis virescens*, and *Pectinophora gossypiella* (Perlak et al. 2001).

The ω -ACTX-Hv1a toxin (Hvt) is a part of the venom of the Australian funnel web spider (*Hadronyche versuta*), which act as a calcium channel antagonist in a heterologous expression system it keeps its biological activity. Finding of an experiment was 100% mortality within 48 h of *Helicoverpa armigera* and *Spodoptera littoralis* larvae feeding on tobacco expressing Hvt (Khan et al. 2006). Biosafety studies for this gene are underway. Another utility of this gene can be protection of crop plants against herbivorous insects. Transformation of *Arabidopsis* vacuolar H^+ pyrophosphatase gene (AVP1) into cotton was carried out recently. Furthermore transformants were found to be resistant against drought and salt stresses and were also having better fiber qualities. This gene has been found to be effective against drought and salt stress in cotton in this study (Pasapula et al. 2010)

5 Implications

5.1 Knowledge-based breeding—proactive decision

Heterosis breeding is an important genetic tool for improving productivities of multiple crop species (Dong et al. 2006). Hybrid maize was first cultivated on a large scale in USA in the 1930s, and later was widely adopted worldwide because of its higher yield potential than the open-pollinated corn varieties. Several reports on interspecific and intraspecific heterosis are available in cotton and it had been the major objective of cotton breeders to use heterosis for lint yield increment (Li 2005). Hybrid vigor for fiber yield and its quality, physiological traits like photosynthetic rate in okra leaf cotton has been claimed (Dong et al. 2006; Zhu et al. 2008); however, yield advantages of cultivating cotton hybrids are not as high as reported for corn.

5.2 Marker-assisted selection for various traits of interest

Numerous QTLs and markers linked with traits have been identified in cotton for utilizing in MAS. For example, QTLs for dense leaf pubescence in cotton (Wright et al. 1999) and QTLs related with drought (Saranga et al. 2001; Paterson et al. 2003; Ullah 2009) have been identified. Similarly, DNA markers associated with nectariless, hairiness, and red leaf color traits were reported (Rahman et al. 2002; Rahman et al. 2002; Ali et al. 2009a, b). Markers linked with a gene which restores male-fertility in upland cotton have further been mapped (Lan et al. 1999). Useful markers linked with two restorer genes (*Rf₁* and *Rf₂*) for developing restorer parental lines were detected in two cotton lines of D₂ genome (Zhang and Stewart 2004; Feng et al. 2005). Markers associated with genes conferring resistance to CLCuD (Rahman et al. 2002; Rahman et al. 2006) were used in probing cotton plants containing the genes in succeeding generations, resulted into the development of two resistant cotton lines NIBGE-2 (Rahman et al. 2008) and NIBGE-115 (Rahman et al. 2008). QTLs for *Fusarium* wilt are a handy tool for conducting MAS (Wang et al. 2009).

DNA markers associated with fiber quality traits are useful for MAS in cotton (Zhang et al. 2003; Asif 2010). A total of nine DNA markers (three SSRs and six random amplified polymorphic DNA (RAPDs)) mapped on one linkage group which were found to be linked with two QTLs for fiber strength. In cotton molecular breeding programs screening for QTLs linked with fiber strength has been successfully conducted with markers (Guo et al. 2003). Chen et al. (2009) further mapped this major fiber strength QTL on Chr-24 (D8) recently.

In MAS for high-quality lint production, markers linked with the QTLs identified in *G. barbadense* are more helpful (Chee et al. 2005b; Mumtaz 2007). Introgression of loci and or genomic regions derived from *GB* into *GH* was tracked by SSRs, resulted in a 2–3-mm increment in fiber length (Mumtaz 2007). Similarly, AFLPs linked with fiber and agronomic traits have also utility in cotton breeding (Jixiang et al. 2007). In another study, a total of 56 QTLs (LOD>3.0) linked with 14 different agronomic and fiber traits and one QTL linked with fiber elongation were found which can be helpful in MAS (Wu et al. 2009).

5.3 Fiber genes—exploring unique genetic pathways

The cotton fiber is an outstanding single-celled model system helpful in studying processes such as cellulose biosynthesis and cell expansion. They differentiate from the protoderm of developing seeds, and are unicellular and unbranched. It has been reported that greater than one-half million quasi-synchronously elongating fibers usually de-

velop in each boll or ovary (Kim and Triplett 2001). Genes involved in fiber development and secondary wall depositions are helpful in understanding many genetic pathways (Haigler et al. 2005).

5.4 Translational genomics

Application of translation of gene functions from a model species to crop plants is the field of “Plant Translational Genomics.” This approach has widely been used for studying the complex traits like tolerance to multiple stresses, plant development, and quality traits of various crop plants. Breeding for these traits is difficult, especially in polyploids (Salentijn et al. 2007). Expanding genomic resources of plants such as *Arabidopsis* (The Arabidopsis genome initiative, 2000), rice (The International Rice Genome Sequencing Program; Yu et al. 2002; Goff et al. 2002), poplar (Tuskan et al. 2006), sorghum (Paterson et al. 2009), and soybean (Schmutz et al. 2010) can provide new opportunities for solving such complex problems. The candidate gene approach is one of the promising methods in this regard (Byrne and McMullen 1996; Pflieger et al. 2001). Similarity between a gene (with known function) in a “model” species (functional candidate genes) with the gene co-localized with the trait-locus (positional candidate genes) in a crop of interest is the basis of this approach. For example, the dwarfing gene *Rht* in wheat is orthologous to genes conferring dwarf mutants, *GAI* in *Arabidopsis* and Dwarf 8 in maize (Peng et al. 1999). This approach has not been much utilized in cotton which can be explored for improving biomass production in developing cotton fibers and cellulose through translational genomic approaches (Chen et al. 2009).

5.5 Genetic resource for other species

Cotton genes can be introduced into other crop species. A 2,062 bases long promoter of *GaMYB2* isolated from *G. arboreum* and fused with a β -glucuronidase (*GUS*) gene for reporting the gene activities in the developing fiber cells and also in the trichomes of other organs such as stems, leaves, and bracts. Trichome-specific promoter was used in *Arabidopsis*. Tobacco plants contain more than one type of trichomes including multicellular simple and glandular secreting trichomes which are different from cotton and *Arabidopsis*—both have non-glandular simple trichomes which are unicellular. In tobacco, the *GaMYB2* promoter has been exclusively involved in limiting *GUS* expression in the glandular cells of glandular secreting trichomes. It was explored after a series of 5'-deletions that a 360-bp fragment upstream of the translation initiation codon was responsible for the gene expression. This comparison of activities of *GaMYB2* promoter specificities in trichomes of

a variety of plant species with different types of trichomes offers an opportunity to further explore plant trichome structure and development (Shangguan et al. 2008).

6 Conclusion

One of the major concerns of stagnation and or decline in cotton productivity worldwide is genetic uniformity among cotton cultivars/germplasm, making cotton yield more vulnerable to biotic and abiotic threats. Some target traits like resistance to drought and heat, insect pests and diseases, and other useful morphological and physiological features have been identified in cultivated diploid cotton species. For introducing useful traits in cultivated cotton varieties, hybridization of multiple species like *G. arboreum*, *Gossypium gossypoides*, *G. herbaceum*, *G. barbadense*, and *Gossypium laxum* with *G. hirsutum* and/or *G. barbadense* followed by raising embryos in culture media may facilitate overcoming cytogenetic barriers. This process would set a stage for widening the genetic base as well as introducing genes conferring traits that are missing in the cultivated species. New experiments like nested association mapping studies, and development of TILLING populations would make productive use of cotton genome sequencing efforts, increasing the utility of the genome sequence in cotton breeding.

Comparative mapping is a powerful tool which provides opportunities to translate information from well-studied genome (s) into applications in less studied genome(s). Enhanced understanding of evolutionary linkage of cotton and *Arabidopsis* has led to the assumptions of organization of respective genes in both genomes. This information facilitates researchers to isolate useful genes from cotton after retrieving functional knowledge from *Arabidopsis*. These translational genomic tools will be further improved after sequencing of cotton genomes. Comparative genomics of cotton and *Arabidopsis* has a remarkable potential to assist in understanding biological pathways. Sequencing of cotton genome can further reveal the minor changes in cotton genome at nucleotide level that affect the functions of genes which further affect the specific economically important traits of cotton.

References

- Abdurakhmonov IY, Kohel RJ, Yu JZ, Pepper AE, Abdullaev AA, Kushanov FN, Salakhutdinov IB, Buriev ZT, Saha S, Scheffler BE, Jenkins JN, Abdulkarimov A (2008) Molecular diversity and association mapping of fiber quality traits in exotic *G. hirsutum* L. germplasm. *Genomics* 92:478–487
- Abelson PH (1998) A third technological revolution. *Science* 279:2019
- Adkisson PL, Niles GA, Walker JK, Bird LS, Scott HB (1982) Controlling cottons insect pests—a new system. *Science* 216:19–22
- Ali I, Kausar A, Rahman M, Zafar Y, Asif M, Ashraf M, Riaz S, Zafar S, Wahid A, Maqsood S, Niaz M, Abbas SQ (2009a) Development of genetic linkage map of leaf hairiness in *Gossypium hirsutum* (cotton) using molecular markers. *Pak J Bot* 41(4):1627–1635
- Ali I, Ashraf M, Rahman M, Zafar Y, Asif M, Kausar A, Riaz S, Niaz M, Wahid A, Abbas SQ (2009b) Development of genetic linkage map of leaf red colour in cotton (*Gossypium hirsutum*) using DNA markers. *Pak J Bot* 41(3):1127–1136
- Arumuganathan K, Earle E (1991) Nuclear DNA content of some important plant species. *Plant Mol Biol Report* 9(3):208–218
- Asad S, Mukhtar Z, Nazir F, Hashmi AJ, Mansoor S, Zafar Y, Arshad M (2008) Silicon carbide whisker-mediated embryogenic callus transformation of cotton (*Gossypium hirsutum* L.) and regeneration of salt tolerant plants. *Mol Biotechnol* 40:161–169
- Asif M (2010) Genomic analysis for quality traits in cotton (*G. hirsutum* L) by DNA fingerprinting technology. B.Z. Univ, Multan Pakistan
- Auld D, Light GG, Fokar M, Bechere E, Allen RD (2009) Mutagenesis system for genetic analysis of *Gossypium*; In genetics and genomics of cotton. Paterson AH (ed), Springer 3: 209–226
- Bolek Y, El-Zik KM, Pepper AE, Bell AA, Magill CW, Thaxton PM, Reddy OUK (2005) Mapping of verticillium wilt resistance genes in cotton. *Plant Sci* 168:1581–1590
- Boulanger J, Pinheiro D (1971) Evolution de la production au nord-est Brasil. *Coton et Fibres Tropicales* 26:319–353
- Bowers JE, Chapman BA, Rong J, Paterson AH (2003) Unraveling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422:433–438
- Brubaker C, Wendel J (1993) Molecular evidence bearing on the specific status of *Gossypium lanceolatum* Todaro. *Genet Resour Crop Evol* 40:165–170
- Brubaker CL, Paterson AH, Wendel JF (1999) Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. *Genome* 42:184–203
- Byrne PF, McMullen MD (1996) Defining genes for agricultural traits: QTL analysis and the candidate gene approach. *Probe* 7:24–27
- Cai Y, Xie Y, Liu J (2010) Glandless seed and glanded plant research in cotton. *Agron Sustain Dev* 30:181–190
- Calhoun DS, Bowman DT, May OL (1997) Pedigrees of upland and pima cotton cultivars released between 1970 and 1995. *Miss Agric For Exp Stn* 1069:53
- Chee P, Draye X, Jiang C-X, Decanini L, Delmonte T, Bredhauer R, Smith CW, Paterson AH (2005a) Molecular dissection of phenotypic variation between *Gossypium hirsutum* and *Gossypium barbadense* (cotton) by a backcross-self approach: III. Fiber length. *Theor Appl Genet* 111:772–781
- Chee P, Draye X, Jiang CX, Decanini L, Delmonte TA, Bredhauer R, Smith CW, Paterson AH (2005b) Molecular dissection of interspecific variation between *Gossypium hirsutum* and *Gossypium barbadense* (cotton) by a backcross-self approach: I. Fiber elongation. *Theor Appl Genet* 111:757–763
- Chen H, Qian N, Guo WZ, Song QP, Li BC, Deng FJ, Dong CG, Zhang TZ (2009) Using three overlapped RILs to dissect genetically clustered QTL for fiber strength on chro.24 in Upland cotton. *Theor Appl Genet* 119:605–612
- Cunnac S, Wilson A, Nuwer J, Kirik A, Baranage G, Mudgett MB (2007) A conserved carboxylesterase is a suppressor of avrbst-elicited resistance in *Arabidopsis*. *Plant Cell* 19:688–705
- Desai A, Chee PW, May OL, Paterson AH (2008) Correspondence of trichome mutations in diploid and tetraploid cottons. *J Hered* 99:182–186
- Dong J, Wu F, Jin Z, Huang Y (2006) Heterosis for yield and some physiological traits in hybrid cotton Cikangza. *Euphytica* 151:71–77
- Draye X, Chee P, Jiang CX, Decanini L, Delmonte TA, Bredhauer R, Smith CW, Paterson AH (2005) Molecular dissection of interspecific variation between *Gossypium hirsutum* and *G.*

- barbadense* (cotton) by a backcross-self approach: II. Fiber fineness. *Theor Appl Genet* 111:764–771
- Endrizzi J, Ramsay G (1979) Monosomes and telosomes for 18 of the 26 chromosomes of *Gossypium hirsutum*. *Can J Genet* 21:531–536
- Feng CD, Stewart JMD, Zhang JF (2005) STS markers linked to the Rf(1) fertility restorer gene of cotton. *Theor Appl Genet* 110:237–243
- Fryxell PA (1963) Morphology of the base of seed hairs of *Gossypium*. I. Cross morphology. *Bot Gaz* 123:196–199
- Fryxell P (1979) The natural history of the cotton tribe. Texas A&M University Press, College Station
- Fryxell PA, Craven LA, Stewart JM (1992) A revision of *Gossypium* sect. *Grandicalyx* (Malvaceae), including the description of six new species. *Syst Bot* 17:91–114
- Goff SA, Ricke D, Lan TH et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. spp. japonica). *Science* 296:92–100
- Gotemare V, Singh P (2004) Use of wild species for cotton improvement in India. *ICAC Rec XXII*:12–14
- Guo WZ, Zhang TZ, Pan JJ, Kohel RJ (1998) Identification of RAPD marker linked with fertility-restoring gene of cytoplasmic male sterile lines in upland cotton. *Chin Sci Bull* 43:52–54
- Guo WZ, Zhang TZ, Shen XL, Yu JZ, Kohel RJ (2003) Development of SCAR marker linked to a major QTL for high fiber strength and its usage in molecular-marker assisted selection in upland cotton. *Crop Sci* 43:2252–2256
- Guo WZ, Cai CP, Wang CB, Han ZG, Song XL, Wang K, Niu XW, Wang C, Lu KY, Shi B, Zhang TZ (2007) A microsatellite-based, gene-rich linkage map reveals genome structure, function and evolution in *Gossypium*. *Genetics* 176:527–541
- Gur A, Zamir D (2004) Unused natural variation can lift yield barriers in plant breeding. *PLoS Biol* 2:1610–1615
- Haigler CH, Zhang DH, Wilkerson CG (2005) Biotechnological improvement of cotton fibre maturity. *Physiol Plant* 124:285–294
- Hao JJ, Yu SX, Dong ZD, Fan SL, Ma QX, Song MZ, Yu JW (2008) Quantitative inheritance of leaf morphological traits in upland cotton. *J Agric Sci* 146:561–569
- Haq I, Asad S, Zafar Y (2005) Bioloistic-mediated transformation of cotton (*Gossypium hirsutum* L.): embryogenic calli as explant. *J Plant Biotechnol* 7(4):211–218
- Helms AB (2000) Yield study report. In: Duggar P, Richder D (eds) *Proc. Beltwide cotton Prod. Conf. San Antonio TX. 4–9 Jan. 2000*. National Cotton Council, Memphis
- Hsu CY, Creech RG, Jenkins JN, Ma DP (1999) Analysis of promoter activity of cotton lipid transfer protein gene LTP6 in transgenic tobacco plants. *Plant Sci* 143:63–70
- Hsu C-Y, Jenkins JN, Saha S, Ma D-P (2005) Transcriptional regulation of the lipid transfer protein gene LTP3 in cotton fibers by a novel MYB protein. *Plant Sci* 168:167–181
- Hulskamp M (2004) Plant trichomes: a model for cell differentiation. *Nat Rev Mol Cell Biol* 5:471–480
- Humphries JA, Walker AR, Timmis JN, Orford SJ (2005) Two WD-repeat genes from cotton are functional homologues of the *Arabidopsis thaliana* TRANSPARENT TESTA GLABRA1 (TTG1) gene. *Plant Mol Biol* 57:67–81
- Jacoby MJ, Falkenhan D, Mader MT, Brininstool G, Wischnitzki E, Platz N, Hudson A, Hulskamp M, Larkin J, Schnittger A (2008) Transcriptional profiling of mature *Arabidopsis* trichomes reveals that NOECK encodes the MIXTA-like transcriptional regulator MYB106. *Plant Physiol* 148:1583–1602
- Jiang CX, Wright RJ, El-Zik K, Paterson AH (1998) Polyploid formation created unique avenues for response to selection in *Gossypium* (cotton). *Proc Natl Acad Sci U S A* 95:4419–4424
- Jiang C, Wright RJ, Woo SS, DelMonte TA, Paterson AH (2000a) QTL analysis of leaf morphology in tetraploid *Gossypium* (cotton). *Theor Appl Genet* 100:409–418
- Jiang C, Chee P, Draye X, Morrell P, Smith C, Paterson A (2000b) Multi-locus interactions restrict gene flow in advanced-generation interspecific populations of polyploid *Gossypium* (cotton). *Evolution* 54:798–814
- Jixiang W, Johnie NJ, McCarty JC, Zhong M, Michael S (2007) AFLP marker associations with agronomic and fiber traits in cotton. *Euphytica* 153:153–163
- Khan SA, Zafar Y, Briddon RW, Malik KA, Mukhtar Z (2006) Spider venom toxin protects plants from insect attack. *Transgenic Res* 15:349–357
- Kim HJ, Triplett BA (2001) Cotton fiber growth in plant and in vitro: models for plant cell elongation and cell wall biogenesis. *Plant Physiol* 127:1361–1366
- Kohel RJ, Richmond TR, Lewis CF (1974) Genetics of flowering response in cotton. VI. Flowering behavior of *Gossypium hirsutum* L. and *G. barbadense* L. hybrids. *Crop Sci* 14:696–699
- Lacape JM, Nguyen TB (2005) Mapping quantitative trait loci associated with leaf and stem pubescence in cotton. *J Hered* 96:441–444
- Lacape JM, Nguyen TB, Thibivilliers S, Bojinov B, Courtois B, Cantrell RG, Burr B, Hau B (2003) A combined RFLP-SSR-AFLP map of tetraploid cotton based on a *Gossypium hirsutum* × *Gossypium barbadense* backcross population. *Genome* 46:612–626
- Lacape JM, Dessauw D, Rajab M, Noyer JL, Hau B (2007) Microsatellite diversity in tetraploid *Gossypium* germplasm: assembling a highly informative genotyping set of cotton SSRs. *Mol Breed* 19:45–58
- Lacape JM, Llewellyn D, Jacobs J, Arioli T, Becker D, Calhoun S, Al-Ghazi Y, Liu S, Palaï O, Georges S, Giband M, de Assunção H, Augusto P, Barroso V, Claverie M, Gawryziak G, Jean J, Vialle M, Viot C (2010) Meta-analysis of cotton fiber quality QTLs across diverse environments in a *Gossypium hirsutum* × *G. barbadense* RIL population. *BMC Plant Biol* 10:132
- Lan T-H, Cook C, Paterson A (1999) Identification of a RAPD marker linked to a male-fertility restoration gene in cotton (*Gossypium hirsutum* L.). *J Agric genomics* 4:1–5
- Larkin JC, Brown ML, Schiefelbein J (2003) How do cells know what they want to be when they grow up? Lessons from epidermal patterning in *Arabidopsis*. *Annu Rev Plant Biol* 54:403–430
- Li JF (2005) Research on Chinese Cotton hybrid vigor utilization. *Jiangxi Cotton* 27(1): 3–7
- Liu HC, Creech RG, Jenkins JN, Ma DP (2000a) Cloning and promoter analysis of the cotton lipid transfer protein gene Ltp3. *Biochimica et Biophysica Acta (BBA) - Mol Cell Biol Lipids* 1487:106–111
- Liu S, Cantrell RG, McCarty JC, Stewart JM (2000b) Simple sequence repeat-based assessment of genetic diversity in cotton race stock accessions. *Crop Sci* 40:1459–1469
- Loguerio LL, Zhang JQ, Wilkins TA (1999) Differential regulation of six novel MYB-domain genes defines two distinct expression patterns in allotetraploid cotton (*Gossypium hirsutum* L.). *Mol Gen Genet* 261:660–671
- Ma X, Zhou B, Lu Y, Guo W, Zhang T (2008) Simple sequence repeat genetic linkage maps of A-genome diploid cotton (*Gossypium arboreum*). *J Integr Plant Biol* 50:491–502
- Marks MD (1997) Molecular genetic analysis of trichome development in *Arabidopsis*. *Annu Rev Plant Physiol* 48:137–163
- Marks MDJP, Wenger E, Gilding R, Jilk, Dixon RA (2009) Transcriptome analysis of *Arabidopsis* wild-type and gl3-sst sim trichomes identifies four additional genes required for trichome development. *Mol Plant* 2(4):803–822
- May OL, Bowman DT, Calhoun DS (1995) Genetic diversity of U.S. upland cotton cultivars released between 1980 and 1990. *Crop Sci* 35:1570–1574
- McCarty J, Jenkins J (1992) Cotton germplasm: characteristics of 79 day-neutral primitive race accessions. *Miss Agric Exp Stn Tech Bull* 184:17

- Mei M, Syed NH, Gao W, Thaxton PM, Smith CW, Stelly DM, Chen ZJ (2004) Genetic mapping and QTL analysis of fiber-related traits in cotton (*Gossypium*). *Theor Appl Genet* 108:280–291
- Mumtaz, H. (2007) Identification of structural and functional genomic markers for fiber quality traits in cotton using interspecific population (*G. hirsutum* x *G. barbadense*). MPhil Thesis, QA Univ Islamabad Pakistan
- Niu C, Hinchliffe DJ, Cantrell RG, Wang CL, Roberts PA, Zhang JF (2007) Identification of molecular markers associated with root-knot nematode resistance in upland cotton. *Crop Sci* 47:951–960
- Niu C, Lister HE, Nguyen B, Wheeler TA, Wright RJ (2008) Resistance to *Thielaviopsis basicola* in the cultivated A genome cotton. *Theor Appl Genet* 117:1313–1323
- Pasapula V, Shen G, Kuppu S, Paez-Valencia J, Mendoza M, Hou P, Chen J, Qiu X, Zhu L, Zhang X, Auld D, Blumwald E, Zhang H, Gaxiola R, Payton P (2010) Expression of an *Arabidopsis vacuolar H⁺-pyrophosphatase* gene (AVP1) in cotton improves drought- and salt tolerance and increases fibre yield in the field conditions. *Plant Biotechnol J* 9:88–99
- Paterson AH, Saranga Y, Menz M, Jiang CX, Wright RJ (2003) QTL analysis of genotype x environment interactions affecting cotton fiber quality. *Theor Appl Genet* 106:384–396
- Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J et al (2009) The Sorghum bicolor genome and the diversification of grasses. *Nature* 457:551–556
- Peng JR, Richards DE, Hartley NM et al (1999) ‘Green revolution’ genes encode mutant gibberellin response modulators. *Nature* 400 (6741):256–261
- Perlak FJ, Oppenhuizen M, Gustafson K, Voth R, Sivasupramaniam S, Heering D, Carey B, Ihrig RA, Roberts JK (2001) Development and commercial use of Bollgard cotton in the USA—early promises versus today’s reality. *The Plant J* 27(6):489–501
- Pflieger S, Lefebvre V, Causse M (2001) The candidate gene approach in plant genetics: a review. *Mol Breed* 7(4):275–291
- Pickersgill B, Barrett SCH, Andrade-Lima D (1975) Wild cotton in northeast Brazil. *Biotropical* 7:42–54
- Plett JM, Mathur J, Regan S (2009) Ethylene receptor ETR2 controls trichome branching by regulating microtubule assembly in *Arabidopsis thaliana*. *J Exp Bot* 60(13):3923–3933
- Quisenberry JE (1977) Inheritance of fiber properties among crosses of Acalas and high plains cultivars of upland cotton. *Crop Sci* 15:202–205
- Quisenberry J, Jordan W, Roark B, Fryrear D (1982) Exotic cottons as genetic sources for drought resistance. *Crop Sci* 21:889–895
- Rahman M, Hussain D, Zafar Y (2002) Estimation of genetic divergence among elite cotton (*Gossypium hirsutum* L.) cultivars/genotypes by DNA fingerprinting technology. *Crop Sci* 42:2137–2144
- Rahman M, Hussain D, Malik TA, Zafar Y (2005) Genetics of resistance to cotton leaf curl disease in *Gossypium hirsutum*. *Plant Pathol* 54:764–772
- Rahman M, Ahmed N, Asif M, Zafar Y (2006) Identification of DNA markers linked with cotton leaf curl disease (CLCD). International Cotton Genome Initiative (ICGI) Workshop, Brasilia Brazil, pp 77–78
- Rahman M, Yasmin T, Tabassum N, Ullah I, Asif M, Zafar Y (2008) Studying the extent of genetic diversity among *Gossypium arboreum* L. genotypes/cultivars using DNA fingerprinting. *Genet Resour Crop Evol* 55:331–339
- Rahman M, Zafar Y, Paterson AH (2009) *Gossypium* DNA markers types, number and uses. In: Paterson AH (ed) *Genomics of cotton*. Springer, Dordrecht
- Rahman M, Asif M, Shaheen T, Tabbasam N, Zafar Y, Paterson AH (2011) Marker-assisted breeding in Higher Plants. In: Eric Lichtfouse (ed) *Sustainable Agriculture Reviews 6; Alternative Farming Systems, Biotechnology, Drought Stress and Ecological Fertilisation*, Springer Publisher Pp 39–76
- Ramsay NA, Glover BJ (2005) MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. *Trends Plant Sci* 10:63–70
- Rayburn ST, Brotton R, Keene E (1999) National cotton variety tests. USDA-ARS, Stoneville
- Rong J, Paterson AH (2009) Comparative genomics of cotton and *Arabidopsis*. *Genetics and Genomics of Cotton*: Paterson AH (ed), Springer 3: 431–449
- Rong J, Abbey C, Bowers JE, Brubaker CL, Chang C, Chee PW, Delmonte TA, Ding XL, Garza JJ, Marler BS, Park C-H, Pierce GJ, Rainey KM, Rastogi VK, Schulze SR, Trolinder NL, Wendel JF, Wilkins TA, Williams-Coplin TD, Wing RA, Wright RJ, Zhao X, Zhu L, Paterson AH (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
- Rong J, Bowers JE, Schulze SR, Waghmare VN, Rogers CJ, Pierce GJ, Zhang H, Estill JC, Paterson AH (2005a) Comparative genomics of *Gossypium* and *Arabidopsis*: unraveling the consequences of both ancient and recent polyploidy. *Genome Res* 15:1198–1210
- Rong J, Pierce G, Waghmare V, Rogers C, Desai A, Chee P, May O, Gannaway J, Wendel J, Wilkins T, Paterson A (2005b) Genetic mapping and comparative analysis of seven mutants related to seed fiber development in cotton. *Theor Appl Genet* 111:1137–1146
- Rong J, Feltus FA, Waghmare VN, Pierce GJ, Chee PW, Draye X, Saranga Y, Wright RJ, Wilkins TA, May OL, Smith CW, Gannaway JR, Wendel JF, Paterson AH (2007) Meta-analysis of polyploid cotton QTLs shows unequal contributions of subgenomes to a complex network of genes and gene clusters implicated in lint fiber development. *Genetics*. doi:10.1534/genetics.107.074518
- Rosenow D, Quisenberry J, Wendt C, Clark L (1983) Drought tolerant sorghum and cotton germplasm. *Agric Water Manag* 7:207–222
- Rungis D, Llewellyn D, Dennis ES, Lyon BR (2002) Investigation of the chromosomal location of the bacterial blight resistance gene present in an Australian cotton (*Gossypium hirsutum* L.) cultivar. *Aust J Agric Res* 53:551–560
- Running MP, Hake S (2001) The role of floral meristems in patterning. *Curr Opin Plant Biol* 4:69–74
- Saha S, Raska D, Stelly DM (2006) Upland (*Gossypium hirsutum* L.) x Hawaiian cotton (*G. tomentosum* Nutt. ex Seem.) F₁ hybrid hypoaneuploid chromosome substitution series. *J Cotton Sci* 10:263–272
- Salentijn EMJ, Pereira A, Angenent GC, Linden CG, Krens F, Smulders MJM, Vosman B (2007) Plant translational genomics: from model species to crops. *Mol Breed* 20:1–13
- Saranga Y, Menz M, Jiang CX, Wright RJ, Yakir D, Paterson AH (2001) Genomic dissection of genotype x environment interactions conferring adaptation of cotton to arid conditions. *Genome Res* 11:1988–1995
- Saranga Y, Jiang CX, Wright RJ, Yakir D, Paterson AH (2004) Genetic dissection of cotton physiological responses to arid conditions and their inter-relationships with productivity. *Plant Cell Environ* 27:263–277
- Schellmann S, Schnittger A, Kirik V, Wada T, Okada K, Beermann A, Thumfahrt J, Jurgens G, Hulskamp M (2002) triptychon and caprice mediate lateral inhibition during trichome and root hair patterning in *Arabidopsis*. *EMBO J* 21:5036–5046
- Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T et al (2010) Genome sequence of the palaeopolyploid soybean. *Nature* 463:178–120
- Shangguan X-X, Xu B, Z-X Yu, Wang L-J, Chen X-Y (2008) Promoter of a cotton fibre MYB gene functional in trichomes of *Arabidopsis* and glandular trichomes of tobacco. *J Exp Bot* 59 (13):3533–3542

- Shen XL, Guo WZ, Zhu XF, Yuan YL, Yu JZ, Kohel RJ, Zhang TZ (2005) Molecular mapping of QTLs for fiber qualities in three diverse lines in Upland cotton using SSR markers. *Mol Breed* 15:169–181
- Shen XL, Van Becelaere G, Kumar P, Davis RF, May OL, Chee P (2006) QTL mapping for resistance to root-knot nematodes in the M-120 RNR Upland cotton line (*Gossypium hirsutum* L.) of the Auburn 623 RNR source. *Theor Appl Genet* 113:1539–1549
- Shen XL, Guo WZ, Lu QX, Zhu XF, Yuan YL, Zhang TZ (2007) Genetic mapping of quantitative trait loci for fiber quality and yield trait by RIL approach in Upland cotton. *Euphytica* 155:371–380
- Silva TMS, Camara CA, Medeiros FD, Oliviera EJ, Argra MF, Harley RM, Giulietti AB (2006) Phaeophytins from *Gossypium mustelinum* Miens ex Watt (Malvaceae). *Biochem Syst Ecol* 34:263–264
- Song XL, Zhang TZ (2007) Identification of quantitative trait loci controlling seed physical and nutrient traits in cotton. *Seed Sci Res* 17:243–251
- Song XL, Guo WZ, Han ZG, Zhang TZ (2005) Quantitative trait loci mapping of leaf morphological traits and chlorophyll content in cultivated tetraploid cotton. *J Integr Plant Biol* 47:1382–1390
- Stewart J. M., 1994. Potential for crop improvement with exotic germplasm and genetic engineering. Challenging the Future: Proceedings of the World Cotton Research Conference-1, Brisbane Australia, G.A. Constable and N.W. Forrester (Eds), CSIRO, Melbourne, Australia, pp. 313–327
- Suo JF, Liang X, Pub L, Zhang YS, Xue YB (2003) Identification of GhMYB109 encoding a R2R3 MYB transcription factor that expressed specifically in fiber initials and elongating fibers of cotton (*Gossypium hirsutum* L.). *Biochim Biophys Acta Mol Cell Biol Lipids* 1630:25–34
- Tester M, Davenport R (2003) Na⁺ tolerance and Na⁺ transport in higher plants. *Ann Bot* 91:503–527
- Tuskan GA, Difazio S, Jansson S, Bohlmann J et al (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 15:1596–1604
- Ullah, I. (2009) Molecular Genetic Studies for Drought Tolerance in Cotton. PhD thesis, Quaid-i-Azam University, Islamabad
- Ulloa M, Meredith WR, Shapley ZW, Kahler AL (2002) Genetic linkage maps from four F2:3 populations and a join maps of *Gossypium hirsutum* L. *Theor Appl Genet* 101:200–208
- Van Esbroeck GA, Bowman DT, Calhoun DS, May OL (1998) Changes in the genetic diversity of cotton in the USA from 1970 to 1995. *Crop Sci* 38:33–37
- Varshney RK, Grosse I, Hahnel U, Siefken R, Prasad M, Stein N, Langridge P, Altschmied L, Graner A (2006) Genetic mapping and BAC assignment of EST-derived SSR markers shows non-uniform distribution of genes in the barley genome. *Theor Appl Genet* 113:239–350
- Wada T, Kurata T, Tominaga R, Koshino-Kimura Y, Tachibana T, Goto K, Marks MD, Shimura Y, Okada K (2002) Role of a positive regulator of root hair development, CAPRICE, in *Arabidopsis* root epidermal cell differentiation. *Development* 129:5409–5419
- Waghmare VN, Rong JK, Rogers CJ, Pierce GJ, Wendel JF, Paterson AH (2005) Genetic mapping of a cross between *Gossypium hirsutum* (cotton) and the Hawaiian endemic, *Gossypium tomentosum*. *Theor Appl Genet* 111:665–676
- Walker AR, Davison PA, Bolognesi-Winfield AC, James CM, Srinivasan N, Blundell TL, Esch JJ, Marks MD, Gray JC (1999) The TRANSPARENT TESTA GLABRA1 locus, which regulates trichome differentiation and anthocyanin biosynthesis in *Arabidopsis*, encodes a WD40 repeat protein. *Plant Cell* 11:1337–1349
- Wang C, Roberts PA (2006) Development of AFLP and derived CAPS markers for root-knot nematode resistance in cotton. *Euphytica* 152:185–196
- Wang S, Wang J-W, Yu N, Li C-H, Luo B, Gou J-Y, Wang L-J, Chen X-Y (2004) Control of plant trichome development by a cotton fiber MYB gene. *Plant Cell* 16:2323–2334
- Wang C, Ulloa M, Roberts PA (2006) Identification and mapping of microsatellite markers linked to a root-knot nematode resistance gene (rkn1) in Acala NemX cotton (*Gossypium hirsutum* L.). *Theor Appl Genet* 112:770–777
- Wang BH, Wu YT, Guo WZ, Zhu XF, Huang NT, Zhang TZ (2007a) QTL analysis and epistasis effects dissection of fiber qualities in an elite cotton hybrid grown in second generation. *Crop Sci* 47:1384–1392
- Wang F, Stewart JM, Zhang J (2007b) Molecular markers linked to the Rf(2) fertility restorer gene in cotton. *Genome* 50:818–824
- Wang HM, Lin ZX, Zhang XL, Chen W, Guo XP, Nie YC, Li YH (2008) Mapping and quantitative trait loci analysis of verticillium wilt resistance genes in cotton. *J Integr Plant Biol* 50:174–182
- Wang P, Su L, Qin L, Hu B, Guo W, Zhang T (2009) Identification and molecular mapping of a *Fusarium* wilt resistant gene in upland cotton. *Theor Appl Genet* 119(4):733–739
- Wendel JF (1989) New world tetraploid cottons contain old-world cytoplasm. *Proc Natl Acad Sci* 86:4132–4136
- Wright R, Thaxton P, El-Zik K, Paterson AH (1998) D-subgenome bias of Xcm resistance genes in tetraploid *Gossypium* (Cotton) suggests that polyploid formation has created novel avenues for evolution. *Genetics* 149:1987–1996
- Wright R, Thaxton P, Paterson AH, El-Zik K (1999) Molecular mapping of genes affecting pubescence of cotton. *J Hered* 90:215–219
- Wu J, Gutierrez OA, Jenkins JN, McCarty JC, Zhu J (2009) Quantitative analysis and QTL mapping for agronomic and fiber traits in an RI population of upland cotton. *Euphytica* 165:231–245
- Yang C, Guo WZ, Li GY, Gao F, Lin SS, Zhang TZ (2008) QTLs mapping for Verticillium wilt resistance at seedling and maturity stages in *Gossypium barbadense* L. *Plant Sci* 174:290–298
- Ynturi P, Jenkins JN, McCarty JC, Gutierrez OA, Saha S (2006) Association of root-knot nematode resistance genes with simple sequence repeat markers on two chromosomes in cotton. *Crop Sci* 46:2670–2674
- Yu J, Nian HS, Wang-Jun et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science* 296:79–92
- Yu JW, Yu SX, Lu CR, Wang W, Fan SL, Song MZ, Lin ZX, Zhang XL, Zhang JF (2007) High-density linkage map of cultivated allotetraploid cotton based on SSR, TRAP, SRAP and AFLP markers. *J Integr Plant Biol* 49:716–724
- Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of Nested Association Mapping. *Genetics* 178:539–551
- Zhang JF, Stewart JM (2004) Identification of molecular markers linked to the fertility restorer genes for CMS-D8 in cotton. *Crop Sci* 44:1209–1217
- Zhang TZ, Yuan YL, Yu J, Guo WZ, Kohel RJ (2003) Molecular tagging of a major QTL for fiber strength in Upland cotton and its marker-assisted selection. *Theor Appl Genet* 106:262–268
- Zhang ZS, Xiao YH, Luo M, Li XB, Luo XY, Hou L, Li DM, Pei Y (2005) Construction of a genetic linkage map and QTL analysis of fibre related traits in upland cotton. *Euphytica* 144:91–99
- Zhao XQ, Xu JL, Zhao M, Lafitte R, Zhu LH, Fu BY, Gao YM, Li ZK (2008) QTLs affecting morph-physiological traits related to drought tolerance detected in overlapping introgression lines of rice (*Oryza sativa* L.). *Plant Sci* 174:618–625
- Zhu W, Liu K, Wang X-D (2008) Heterosis in yield, fiber quality, and photosynthesis of okra leaf oriented hybrid cotton (*Gossypium hirsutum* L.). *Euphytica* 164(1):283–291
- Zuo K, Sun J, Zhang X, Nie Y, Liu J, Feng C (2000) Constructing a linkage map of upland cotton (*Gossypium hirsutum* L.) using RFLP, RAPD and SSR makers. *J Huazhong Agric Univ* 19:190–193