We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

3,800
Open access books available

116,000
International authors and editors

120M
Downloads

154
Countries delivered to

TOP 1%
Our authors are among the most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
Genetic Diversity in Gossypium genus

Ibrokhim Y. Abdurakhmonov et al*
Center of Genomic Technologies, Institute of Genetics and Plant Experimental Biology
Academy of Sciences of Uzbekistan, Yuqori Yuz, Qibray Region, Tashkent
Uzbekistan

1. Introduction
Cotton (Gossypium spp.) is the unique, most important natural fiber crop in the world that brings significant economic income, with an annual average ranging from $27 – 29 billion worldwide from lint fiber production (Campbell et al., 2010). The worldwide economic impact of the cotton industry is estimated at ~$500 billion/yr with an annual utilization of ~115-million bales or ~27-million metric tons (MT) of cotton fiber (Chen et al., 2007). In 2011 and 2012, global cotton production is projected to increase 8% (to 26.9 million MT). This will be the largest crop since 2004 and 2005 (International Cotton Advisory Committee [ICAC], 2011).

Cotton is also a significant food source for humans and livestock (Sunilkumar et al., 2006). Cotton fiber production and its export, being one of the main economic resources, annually brings an average of ~$0.9 to 1.2 billion economic income for Uzbekistan (Abdurakhmonov, 2007) that represented 22% of all Uzbek exports from 2001-2003 (Campbell et al., 2010). The economic income from cotton production accounts for roughly 11% of the Uzbekistan's GDP in 2009 (http://www.state.gov/r/pa/ei/bgn/2924.htm, verified on September 15, 2011).

The level of genetic diversity of crop species is an essential element of sustainable crop production in agriculture, including cotton. The amplitude of genetic diversity of Gossypium species is exclusively wide, encompassing wide geographic and ecological niches. It is conserved in situ at centers for cotton origin (Ulloa et al., 2006) and preserved ex situ within worldwide cotton germplasm collections and materials of breeding programs. Cotton
productivity and the future of cotton breeding efforts tightly depend on 1) the level of the genetic diversity of cotton gene pools and 2) its effective exploitation in cotton breeding programs. Elucidating the details of genetic diversity is also very important to determine timeframe of cotton agronomy, develop a strategy for genetic gains in breeding, and conserve existing gene pools of cotton.

During past decades, because of advances in molecular marker technology, there have been extensive efforts to explore the molecular genetic diversity levels in various cotton gene pools and genomic groups, varietal and breeding collections, and specific germplasm resources. These efforts reinforced a serious concern about the narrow genetic base of cultivated cotton germplasm, which has obviously been associated with a “genetic bottleneck” occurred during historic cotton domestication process (Iqbal et al., 2001). A narrow genetic base of cultivated germplasm was one of the major factors causing the recent cotton yield and quality declines (Esbroeck & Bowman, 1998; Paterson et al., 2004). These declines, however, are largely due to challenges and the lack of innovative tools to effectively exploit genetic diversity of Gossypium species. The most effective utilization of genetic diversity of Gossypium species further requires modern genomics technologies that help to reveal the molecular basis of genetic variations of agronomic importance. Sequencing the cotton genome(s) (Chen et al., 2007;) is a pivotal step that will facilitate the fine-scale mapping and better utilization of functionally significant variations in cotton gene pools (Abdurakhmonov, 2007). Once exploited effectively, these wide ranges of genetic diversity of the genus, in particular reservoir of potentially underutilized genetic diversity in exotic wild cotton germplasm, are the ‘golden’ resources to improve cotton cultivars and solve many fundamental problems associated with fiber quality, resistance to insects and pathogens and tolerance to abiotic stresses (Abdurakhmonov, 2007). In this chapter, we describe cotton germplasm resources, the amplitude of morphobiological and agronomic diversity of Gossypium genus and review efforts on molecular genetic diversity of cotton gene pools as well as highlight examples, challenges and perspectives of exploiting genetic diversity in cotton.

2. Description of cotton gene pools and worldwide germplasm collections

Although wild cottons (Gossypium spp) are perennial shrubs and trees, the domesticated cottons are tropic and sub-tropic annual crops cultivated since prehistoric times of the development of human civilization. The Gossypium genus of the Malvaceae family contains more than 45 diploid species and 5 allotetraploid species (Fryxell et al., 1992; Percival et al., 1999; Ulloa et al., 2007). These species are grouped into nine genomic types ($x = 2n = 26$, or $n = 13$) with designations: AD, A, B, C, D, E, F, G, and K (Percival et al., 1999). The species are largely spread throughout the diverse geographic regions of the world. Based on the usage of these Gossypium species in cotton breeding and their genetic hybridization properties, they can be grouped into 1) primary gene pool, which includes the two species from the New World, G. hirsutum L. and G. barbadense L., as well as remaining three wild tetraploid species, G. tomentosum Nuttall ex Seemann, G. mustelinum Miers ex Watt and G. darwinii Watt; 2) secondary gene pool, including A, B, D and F genome diploid cotton species; 3) tertiary gene pool, including C, E, G, K genome Gossypium species (Stelly et al., 2007; Campbell et al., 2010).
Diploid cottons, referred as Old World cottons, are classified into eight (A-G to K) cytogenetically defined genome groups that have African/Asian, American, and Australian origin (Endrizzi et al., 1985). Two of these Old World cottons from Asian origin, G. arboreum L. and G. herbaceum L., with a spinnable seed fiber, were originally cultivated in Asian continent. Today, Old World cultivated cottons remain primarily for non-industrial consumption in India and adjacent Asian countries.

The New World diploid *Gossypium* comprises of 14 (one undescribed taxon US-72) D genome species (Ulloa et al., 2006; Alvarez and Wendel, 2006; Feng et al., 2011). Taxonomically, these species are recognized as the *Houzingenia* subgenus (Fryxell, 1979, 1992). Twelve of the 14 species of this group are distributed in Mexico and extending northward into Arizona. Five species are adapted to the desert environments of Baja California [G. armourianum Kearney (D₂₋₁), G. harknessii Brandegee (D₂₋₂), and G. davidsonii Kellogg (D₃₋₃)] and NW mainland Mexico [G. turneri Fryxell (D₁₀) and G. thurberi Todaro (D₁)]. An additional seven species [G. sp. US-72, G. aridum (Rose & Standley) Skovsted (D₂), G. lobatum Gentry (D₁₁), G. laxum Phillips (D₄), G. schwendimannii Fryx. & Koch (D₁₁), G. gossypioideis (Ullbrich) Standley (D₄₁), and G. trilobum (Mociño & Sessé ex DC.) Skovsted (D₃₉)] are located in the Pacific coast states of Mexico and, with the exception of the last species, are arborescent in growth habit (Ulloa et al., 2006). The other two species with disjunct distributions, G. raimondii Ullbrich (D₁₃) is endemic to Peru, while G. klotzschianum Andersson (D₂₋₄) is found in the Galápagos Islands. The D-genome species (subgenus *Houzingenia*) are classified into six sections: Section *Houzingenia* Fryxell (D₁₋₃ and D₃₋₃); Section Integrifolia Todaro (D₃₋₄ and D₃₋₃); Section Caducibracteolata Mauer (D₂₋₁, D₂₋₂, and D₁₀); Section Erioxylum Rose & Standley (US-72, D₁, D₇, D₉, and D₁₁); Section Selera (Ullbrich) Fryxell (D₂₀); and Section Austroamericana Fryxell (D₃₁) (Percival et al., 1999).

Until recently, evaluation of the New World D-genome species of *Gossypium*, especially Section *Houzingenia* and Section *Erioxylum*, has been limited by the lack of resource material for *ex situ* evaluation. In recent years, the United States Department of Agriculture and the Mexican Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias (INIFAP) have sponsored joint *Gossypium* germplasm collection trips by U.S. and Mexican cotton scientists (Ulloa et al., 2006; Feng et al., 2011). As a result of these efforts, a significant number of additional *Gossypium* accessions of the subgenus *Houzingenia* from various parts of Mexico are now available for evaluation, including several accessions of each of the arborescent species (Ulloa et al., 2006). Although none of these diploid species produces cotton fibers, the D genome is one of the parental lineages of the modern allotetraploid cultivated cottons, Upland and Pima (Ulloa, 2009). Studying these D genome species is the first critical step to fulfill the pressing need to document the *in situ* conservation, to assess the genetic diversity in *Gossypium* species for the preservation of the D genome species, and to facilitate their use for cotton improvement. *In situ* conservation of some of these species is threatened by population growth and industrialized agriculture. These *Gossypium* species are donors of important genes for cotton improvement (Ulloa et al., 2006).

Hybridization between A-genome (Old World cottons) and D-genome (New World cottons) diploids and subsequent polyploidization about 1.5 million years ago created the five AD allotetraploid lineages belonging to the primary gene pool that are indigenous to America and Hawaii (Phillips, 1964; Wendel & Albert, 1992; Adams et al., 2004). These New World allotetraploid cottons include the commercially important species, *G. hirsutum* and *G.
barbadense, which are extensively cultivated worldwide (Abdurakhmonov, 2007; Campbell et al., 2010).

G. hirsutum (also called Acala or Upland, short stapled, Mocó, and Cambodia cotton) is the most widely cultivated (90%) and industrial cotton among all Gossypium species. It includes the Upland cotton cultivars and other early maturing, annually grown herbal bushes. The center of origin for G. hirsutum is Mesoamerica (Mexico and Guatemala), but it spread throughout Central America and Caribbean. According to archaeobotanical findings, G. hirsutum probably was domesticated originally within the Southern end of Mesoamerican gene pool (Wendel, 1995; Brubaker et al., 1999). Consequently, two centers of genetic diversity exist within G. hirsutum: Southern Mexico-Guatemala and Caribbean (Brubaker et al., 1999); Mexico-Guatemala gene pool is considered the site of original domestication and primary center of diversity. Within this range, G. hirsutum exhibits diverse types of morphological forms, including wild, primitive to domesticated accessions. According to Mauer (1954), there are four groups of sub-species of G. hirsutum: (1) G. hirsutum ssp. mexicanum, (2) G. hirsutum ssp. paniculatum, (3) G. hirsutum ssp. punctatum, and (4) G. hirsutum ssp. euhirsutum (domesticated cultivars). These four groups of sub-species include within themselves a number of wild landraces and primitive predominated forms such as yucatanense, richmondi, punctatum, latifolium, palmeri, morilli, purpurascens and their accessions as well as a number of domesticated variety accessions from 80 different cotton growing countries worldwide (Sunilkumar et al., 2006; Lacape et al., 2007; Abdurakhmonov, 2007).

G. barbadense (also called as long staple fibered Pima, Sea Island or Egyptian cotton), accounting for about 9% of world cotton production, was originally cultivated in coastal islands and lowland of the USA and became known as Sea Island cotton. Sea Island cottons, then, were introduced into Nile Valley of Egypt and widely grown as Egyptian cotton to produce long staple fine fibers (Abdalla et al., 2001). The wide-distribution of G. barbadense included mostly South America, southern Mesoamerica and the Caribbean basin (Fryxell, 1979). G. barbadense can be divided into two botanical races brasilense (with kidney-seed trait) and barbadense (with nonaggregated seeds) that both widely present as semi-domesticated forms in Brazil (de Almeida et al., 2009). The brasilense race, considered to have been domesticated in the Amazonian basin (de Almeida et al., 2009) is considered a locally domesticated form for G. barbadense cotton (Brubaker et al., 1999; de Almeida et al., 2009).

The other three AD tetraploid species of cotton, G. mustelinum with specific distribution in the Northeast Brazil (Wendel et al., 1994), G. darwinii endemic to Galapagos Islands (Wendel & Percy, 1990), and G. tomentosum Nutall ex Seemann endemic to Hawaiian Islands (Dejoode and Wendel, 1992; Hawkins et al., 2005), are truly wild species (Westengen et al., 2005).

The main ex situ cotton germplasm collections are in the US, France, China, India, Russia, Uzbekistan, Brazil, and Australia. Although there are a few other cotton germplasm collections present in other countries of the world, these eight countries represent the majority of the world’s cotton germplasm resources. Each country has a germplasm storage and conservation program in place (Campbell et al., 2010). The history of collecting an initial cotton germplasm through the specific expeditions of cotton scientists to the centers of Gossypium origins are well described by Ulloa et al. (2006) that were the basis, perhaps, for the majority of the current cotton germplasm collections worldwide. Consequently, to protect the world-wide economic value of cotton and cotton byproducts, cotton germplasm collections worldwide were enriched with numerous cotton germplasm accessions and
breeding materials/lines as source of the genetic diversity through continuous research efforts of specific cotton breeding programs and mutual germplasm exchange over the last 100 years (Abdurakhmonov, 2007; Campbell et al., 2010).

The brief descriptions for some of worldwide cotton germplasm collections were highlighted in several documents by Abdurakhmonov (2007), Chen et al. (2007), Stelly et al. (2007), Ibragimov et al. (2008), Wallace et al. (2009) and Campbell et al. (2010). In particular, a recent report of cotton researchers published in Crop science journal (Campbell et al., 2010) has widely described the current status of global cotton germplasm resources. Campbell et al. (2010) provided information regarding: 1) members of the collection, 2) maintenance and storage procedures, 3) seed request and disbursement, 4) funding apparatus and staffing, 5) characterization methodology, 6) data management, and 7) past and present explorations.

The contents and distribution of cotton germplasm accessions across the eight collections is summarized by Campbell et al. (2010), so we will not review the details of each collection to avoid redundancy, but rather found appropriate to list brief information in regards to the overall content and specificity of these world cotton collections. Based on a number of preserved cotton accessions in the collection, the eight major world collections can be positioned as follows: Uzbekistan (18971 accessions), India (10469 accessions), USA (10318 accessions), China (8837 accessions), Russia (6276 accessions), Brazil (4296 accessions), CIRAD (France; 3070 accessions) and Australia (1711 accessions). The main content of these collections consists of accessions for two cultivated cotton species, G. hirsutum and G. barbadense. Uzbekistan (2680 accessions), India (2283 accessions) and USA (1923 accessions) collections are the richest ones to maintain a great number of accessions for Asian diploid cottons, G. herbaceum and G. arboreum belonging to the secondary gene pool. If the collection of wild species belonging to primary, secondary and tertiary gene pools are considered Brazil (889 accessions), USA (509 accessions) and CIRAD (295 accessions) are the richest cotton collections in the world.

3. Spectra of morphological and agronomic diversity in cotton

The amplitude of genetic diversity of cotton (Gossypium spp), including all its morphological, physiological and agronomic properties, is exclusively wide (Mauer, 1954). There is a great deal of genetic diversity in the Gossypium genus with characteristics such as plant architecture, stem pubescence and color, leaf plate shape, flower color, pollen color, boll shape, fiber quality, yield potential, early maturity, photoperiod dependency, and resistance to multi-adversity environmental stresses that are important for the applied breeding of cotton. The glimpse of genetic diversity on some morphological traits is demonstrated in Figs. 1 and 2.

Besides morphological diversity in Gossypium genus, representatives of different genomic groups have diverse characteristics in many agronomically useful traits. Considering only G. hirsutum accessions, exotic and cultivar germplasm represent a wide range of genetic diversity in yield and fiber quality parameters. For example, in the analyses of ~1000 G. hirsutum exotic and cultivated accessions in the two different environments, Mexico and Uzbekistan, we found a wide range of useful agronomic diversities (Abdurakhmonov et al., 2004, 2006, 2008, 2009). In one or two environments, the cotton boll mass varies in a range of
1-9 grams per boll, 1000 seed mass varies in a range of 50-170 grams, the lint percentage varies in a range of 0-45%, Micronaire varies in a range of 3-7 mic, the fiber length varies in a range of 1-1.28 inch, and fiber strength varies in a range of 26-36 g/tex. There was also a wide range of variation in photoperiodic flowering (day neutral, weak to strong photoperiodic dependency) and maturity (Abdurakhmonov, 2007). This wide phenotypic diversity of cotton shows the extensive plasticity of cotton plants and potential of their wide utilization in the breeding programs as an initial material.

4. Some examples of exploiting genetic diversity through traditional breeding

Above mentioned genetic diversity, preserved in germplasm collections worldwide, are the golden resources to genetically improve the cotton cultivars. There are numerous examples on the utilization of such genetic variations in solving many fundamental problems in cotton breeding and production (Abdurakhmonov, 2007). For instance, the exploration for genetic diversity for Verticillium wilt fungi from the exotic G. hirsutum spp mexicanum var nervosum germplasm and its on-time mobilization into the elite cultivars solved wilt epidemics in 1960's and saved Uzbekistan's cotton production, and so the economy of the country (Abduillaev et al., 2009). As a result, the wilt resistant variety series named as “Tashkent” were developed (Abdukarimov et al., 2003; Abdurakhmonov, 2007). Later, salt tolerant genotype AN-Boyovut-2 was selected from Tashkent cultivar biotypes demonstrating a continuation of a ‘genetic diversity imprint’ introgressed from the wild landrace stock (Abdukarimov et al., 2003). This is one of the success stories on exploiting genetic diversity and its impact from the single landrace stock germplasm, G. hirsutum spp. mexicanum (Abdurakhmonov, 2007). A number of other examples on the creation of natural defoliation, disease and pest resistance, tolerance to multi-adversity stresses, improved seed oil content and fiber quality parameters, utilizing the exotic germplasm genetic diversity in Uzbekistan have been well documented (Abdukarimov et al., 2003; Abdurakhmonov et al., 2005, 2007).

Successful photoperiodic conversion program in cotton was developed to mobilize day-neutral genes into the primitive accessions of G. hirsutum. Day-neutral genes were introgressed into 97 primitive cotton accessions by a large backcrossing effort (McCarty et al., 1979; McCarty & Jenkins, 1993, Liu et al., 2000). This converted cotton germplasm is an important reservoir for potential genetic diversity and can be used as a source to introgress genes into breeding germplasm (Abdurakhmonov, 2007).

Similarly, using genetic diversity existing in Gossypium genus, reniform nematode resistance, which is one of the high cost ($100 million/year) problems in US cotton production, was addressed. Scientists succeeded in introgressing high resistance to the nematode from G. longicalyx into G. hirsutum through genetic bridge crossing of two trispecies hybrids of G. hirsutum, G. longicalyx, and either G. armouriannum or G. herbaceum (Robinson et al., 2007). Later, a gene of interest was mapped (Dighe et al., 2010). Resistance to root-knot nematode was also solved with the use of genetic diversity in Gossypium genomes (Roberts & Ulloa, 2010). Additionally, Hinze et al. (2011) developed four diverse populations based on US germplasm collection that helped to utilize a large amount of ‘still underutilized’ genetic variability in cotton breeding that should be useful in sustainable cotton production with superior quality. There are many other examples recorded in different cotton breeding programs, but we limit this section with above examples and move to address the challenges behind these success stories and future perspectives in this direction.
Fig. 1. Morphological trait diversity: (A)-pollen color, petal color and spot; (B)-matured bolls in diploid species, and (C)-matured bolls in tetraploid species.
Fig. 2. Diversity on several morphological traits in exotic and cultivar germplasm of *G. hirsutum* accessions from Uzbekistan cotton collection (Abdurakhmonov et al., 2004, 2006).
5. Challenges and perspectives of exploiting diversity of different gene pools

The introduction of genetic diversity into elite cotton germplasm is difficult and the breeding process is slow. When breeders use new and exotic germplasm sources, which possess desirable genes for crop trait improvements, large blocks of undesirable genes are also introgressed during the recombination between the two parental lines (linkage drag). This linkage drag has limited the use of such germplasm. Therefore, the utilization of useful genetic diversity of the wild germplasm using traditional breeding efforts is challenging due to: 1) hybridization issues between various cotton genomes, 2) sterility issues of interspecific multi-genome hybrids, 3) segregation distortion, 4) photoperiodic flowering of wild cottons and 5) long timescale (10-12 years of efforts) required for successful introgression and recovering superior quality homozygous genotypes using traditional breeding approaches (Abdurakhmonov, 2007). This underlies necessity for the development of new innovative genomics approaches to support and accelerate the traditional efforts of exploiting the genetic diversity in cotton breeding. Continuing the introduction of genetic diversity into cultivated plants is important for reducing crop vulnerability and improving important traits such as yield, fiber quality traits, and disease and pest resistance of the cotton crop.

The most effective utilization of the genetic diversity of *Gossypium* species further requires (1) characterization of candidate gene(s) underlying the phenotypic and agronomic diversities based on genomic information in other species, (2) estimation of molecular diversity, genetic distances, genealogy and phylogeny of gene pools and germplasm groups, (3) acceleration of linkage mapping and marker-assisted selection, (4) development of efficient cotton transgenomics, and (5) sequencing cotton genome(s) (Abdurakhmonov, 2007). Furthermore, (6) it is very important to characterize and describe the existing cotton germplasm collections for both phenotypic and genomic diversity. Consequently, (7) incorporation of information into electronic web-based cotton databases such as cotton DB (http://cottondb.org), Cotton Portal (http://gossypium.info), and the Cotton Diversity Database (http://cotton.agtec.uga.edu; Gingle et al., 2006) as well as further improvement of data management tools are pivotal to facilitate an effective exploitation of the genetic diversity of cotton in the future. Cotton germplasm exchange (8) among collections and research groups is also an imperative part toward this goal (Abdurakhmonov, 2007).

6. Characterization of molecular genetic diversity in *Gossypium* genus

Molecular diversity using protein and DNA marker technologies has extensively been studied for accessions from primary and secondary gene pools. Molecular genetic diversity of tertiary gene pool cotton species is poorly explored using molecular marker technology.

6.1 Molecular diversity within primary gene pool

6.1.1 Upland germplasm

As mentioned above, application of modern molecular marker technologies, such as isozymes (Wendel & Percy, 1990; Wendel et al., 1992), random amplified polymorphic DNAs - RAPDs (Multani & Lyyon, 1995; Tatineni et al., 1996; Iqbal et al., 1997; Mahmood et al., 2009; Chaudhary et al., 2010), restricted fragment length polymorphisms - RFLPs (Wendel & Brubaker, 1993), amplified fragment length polymorphisms - AFLPs (Pilay & Myers, 1999; Abdalla et al., 2001; Iqbal et al., 2001; Rana et al., 2005; Lukonge et al., 2007) and
Simple Sequence Repeats – SSRs (Liu et al., 2000, Gutierrez et al., 2002; Rungis et al., 2005; Zhang et al., 2005a; Bertini et al., 2006; Zhang et al., 2011a; Kalivas et al., 2011) generally revealed a low level of genetic diversity within Upland cultivars. There were little variations in estimation of molecular diversity among Upland cultivars (G. hirsutum); however, in general, the genetic distance reported for Upland cultivars was in the range of 0.01-0.28 (Abdurakhmonov, 2007).

Recently, we analyzed a large number of G. hirsutum variety and exotic accessions from Uzbek cotton germplasm collection (Fig.2) with SSR markers (Abdurakhmonov et al., 2008, 2009). Analysis of a large number of G. hirsutum accessions from exotic germplasm and diverse ecotypes/breeding programs with SSR markers confirmed the narrow genetic base of Upland cotton cultivar germplasm pool (with the genetic distance (GD) range of 0.005-0.26) and provided an additional evidence for the occurrence of a genetic ‘bottleneck’ during domestication events of the Upland cultivars at molecular level (Iqbal et al., 2001). Molecular diversity analysis of germplasm accessions using principal component analysis (PCA) suggested that germplasm resources could be broadly grouped into three large clusters (Fig.3) of exotic (1), USA-type (2) and Uzbekistan (3). First three eigenvalues of PCA analysis accounted for a ~52% variation and demonstrated existence of wide genetic diversity within the exotic germplasm, including germplasm accessions from Mexican and African origin (GD=0.02-0.50; Fig.3). We recorded a plenty of private SSR alleles within each group of accessions, specific to the germplasm groups, breeding ecotypes or exotic accessions.

A wider genetic diversity in the land race stocks of G. hirsutum was reported by previous studies (Liu et al., 2000, Lacape et al., 2007), suggesting the existence of sufficient genetic diversity in the exotic germplasm for future breeding programs. Rana et al. (2005) also reported a wider genetic diversity (30-87%) within G. hirsutum breeding lines using AFLP markers. Some recent studies have reported a relatively higher genetic diversity with an average genetic distance of up to ~37-77% in G. hirsutum cultivars, based on the analysis of specific germplasm resources from Brazil (Bertini et al., 2006), Pakistan (Khan et al., 2009; Azamat & Khan, 2010), India (Chaudhary et al., 2010) and China (Liu et al., 2011; Zhang et al., 2011a) breeding programs. Results of these studies were inferred from SSR or combination of a SSR and/or RAPD marker polymorphisms.

Similarly, using SSR and RAPD markers, Sapkal et al. (2011) reported moderately high level of genetic diversity (up to 57%) for 91 Upland cotton accessions with genetic male sterility maintainer and restorer properties. This suggested the existence of useful genetic diversity both in exotic and breeding line resources, useful to broaden the genetic base of Upland cotton cultivars. There is a need for evaluation of molecular genetic diversity level (Zhang et al., 2011a) and its effective exploitation in breeding programs that will address current concerns on narrowness of genetic base of widely grown Upland cotton cultivars (Hinze et al., 2011).

6.1.2 Sea Island germplasm

The molecular genetic diversity within G. barbadense germplasm accessions was also studied using molecular markers such as allozymes (Wendel & Percy, 1990) and AFLPs (Abdalla et al., 2001; Westengen et al., 2005). These studies revealed a narrow genetic base within G. barbadense accessions with a genetic distance of 7-11% (Abdalla et al., 2001; Westengen et al., 2005)
Fig. 3. Principal coordinate analysis of Upland cotton (G. hirsutum) accessions from Uzbek cotton germplasm collection analyzed with SSR markers. Two (A) and three (B) dimensional view for accessions from Africa (AF), Afghanistan (AFG), Bulgaria (BUL), China (CN), Czechoslovakia (CZ), India (IN), Korea (KOR), Mexico (MEX), Pakistan (PAK), Syria (SYR), Ukraine (UA), United States (US), Uzbekistan (UZB), Yugoslavia (YU), and others (Turkey, Iraq and Azerbaijan).

as was observed within the Upland cotton germplasm. In contrast, Boopathi et al. (2008) have identified highly diverse pairs of G. barbadense accessions using SSR marker analysis, which is useful for breeding of high quality Pima type cotton cultivars. Recently, de Almeida et al. (2009) have studied the molecular diversity level of G. barbadense populations...
in situ preserved in the two states of Brazil, Amapá and Para. The genetic analysis using SSR markers of plant populations in these two states revealed 1) high homozygosity in each genotype tested, 2) high total genetic diversity ($H_e=39\%$) in *G. barbadense* populations studied and 3) high level of population differentiation ($F_{st}=36\%$) between cotton plants from these two Brazilian states. Results suggested the existence of noticeable genetic diversity preserved in *in situ* populations of *G. barbadense* in Brazil that should be further maintained within an *ex situ* germplasm collection to guarantee its long term preservation (de Almeida et al., 2009). Similarly, there is useful genetic diversity in *ex situ* preserved *G. barbadense* germplasm collections worldwide. For instance, the molecular diversity analysis of *G. barbadense* accessions using SSR markers revealed that moderately higher genetic diversity (up to 34%) exists within former USSR (that includes collections of Uzbekistan and Russia), China, USA, and Egypt germplasm collections (Wu et al., 2010). In that, USSR collection demonstrated the extraordinary genetic diversity compared with other collections whereas Egyptian collection had the least genetic diversity.

6.1.3 Wild allotetraploid germplasm

The molecular diversity revealed by AFLP markers was low within *G. tomentosum* germplasm with a genetic distance range of 2-11% (Hawkins et al., 2005). However, recent efforts on the characterization of genetic diversity level of three *in situ* preserved *G. mustelinum* population from Brazil using SSR markers suggested 1) high level of homozygosity within each population studied and 2) existence of high level of total genetic differentiations (58.5%) between them, which is due to geographic isolations and genetic founder effects (Barroso et al., 2010). Wendel & Percy (1990) analyzed 58 *G. darwinii* accessions from six islands using 17 isozyme markers encoded by 59 genetic loci and identified high genetic diversity level within its accessions and relationships with *G. barbadense* and *G. hirsutum* genomes. This classical study suggested that *G. darwinii* is closely related to *G. barbadense* despite having gene flow imprints from *G. hirsutum*; however, *G. darwinii* has a large number of unique alleles to be considered a distinct genome (Wendel & Percy, 1990).

6.2 Molecular diversity within secondary gene pool

The genomic diversity of the A-genome diploid cottons has also been studied using molecular marker technology (Liu et al., 2006; Guo et al., 2006; Kebede et al., 2007; Rahman et al., 2008; Kantartzis et al., 2009; Patel et al., 2009; Azamat & Khan, 2010). The genetic distance within 39 *G. arboreum* L (A$_2$A$_2$-genome) accessions, analyzed with SSR markers, ranged from 0.13-0.42 (Liu et al., 2006) demonstrating the existence of wider genomic diversity in the A-genome diploids compared to the Upland cultivar germplasm. Kebede et al. (2007) reported, however, moderate level of genetic diversity within each A$_1$ and A$_2$-genome cottons that ranged from 0.03-0.20 with an average of 0.11 within *G. herbaceum* and 0.02-0.18 with an average of 0.11 for *G. arboreum* (A$_3$). The overall genetic distance between A$_1$ and A$_2$ genomes was up to 36-38% (Kebede et al., 2007; Mahmood et al., 2010). In fact, *G. arboreum* arose from the primitive perennial form of *G. herbaceum* spread in India and there is a single reciprocal chromosomal translocation in *G. arboreum* genome compared to *G. herbaceum* (Guo et al., 2006). Molecular diversity revealed by SSR markers was higher within *G. arboreum* accessions (an average of 25%) compared to *G. herbaceum* accessions (an average of 4%; Patel et al., 2009), suggesting differences in two closely related cotton genome
germplasm resources. This is an interesting finding but is in contrast to the report by Kebede et al. (2007) where an average genetic diversity within A1 and A2 genome accessions was equal.

Rahman et al. (2008) studied 32 G. arboresum accessions specific to Pakistan with RAPD markers and found up to 53% genetic diversity between studied accessions with very narrow diversity within cultivated G. arboresum accessions compared to non-cultivated ones. Analyzing 96 G. arboresum accessions with SSR markers, Kantartzzi et al. (2009) reported that genetic distance within these geographically diverse A2 genome accessions ranged up to 51%. In a more recent study, Azamat & Khan (2010) also reported wider genetic diversity in G. arboresum cultivar germplasm revealed by RAPD (GD=0.371) and SSR markers (GD=0.41). Although variable genetic distance estimates are presented, these reports collectively suggest that A genome representatives of secondary gene pool have sufficient molecular diversity useful for breeding programs.

Studying a large number of accessions for D genome cotton such as G. aridum (D4), G. davidsonii (Dx4), G. klotzschianum (Dx4), G. laxum (D6), G. lobatum (D7), G. schwendimanii (D14) with AFLP markers Alvarez & Wendel (2006) have reported 7 to 54% genetic diversity among D-genome accessions studied. A wider range of genetic diversity was observed among 12 D-genome diploid cottons with the genetic similarity of 0.08-0.94 (Guo et al., 2007a), suggesting existence of diverse variations in D-genome cotton germplasm useful for breeding programs. Recently, Feng et al. (2011) have studied 33 arborescent D-genome accessions, including 23 accessions of G. aridum with RAPD and AFLP markers. They found high molecular diversity among accessions studied, varying from 32% to 84%. This study suggests for continual efforts to study these D-genome American Gossypium species (subsection Erioxylum) to resolve genetically distant geographical ecotypes useful for cotton improvement (Feng et al., 2011).

6.3 Molecular diversity within tertiary gene pool

There is a limited information on molecular diversity estimates for tertiary germplasm pool accessions, including C, E, G and K-genomic species. Recently, Tiwari & Stewart (2008) reported AFLP marker-based molecular diversity analysis results for 57 accessions of C- and G-genome species, including G. australe F. Mueller (G), G. nelsonii Fryxell (G), G. prokanov Prokhanov (G3) and G. sturtianum J.H. Willis (C). Results showed that within G. australe accessions, the pairwise mean genetic distance was in a range of 3-15%, suggesting narrow genetic diversity within G. australe accessions that could be due to relatively recent seed dispersal over large growing area of this species (Tiwari & Stewart, 2008). However, there was moderately high molecular diversity between G. australe and G. nelsonii accessions, ranging from ~17-31%. Higher molecular diversity of up to ~43% was found between G. australe and G. bickii accessions. The genetic distance between G. bickii and G. nelsonii varied from 25% to 35% and as expected C1-genome accessions were most distantly related ones to these three G-genome species (Tiwari & Stewart, 2008). There is no report on molecular diversity studies on other representatives of tertiary gene pool.

6.4 Molecular diversity among cotton gene pools

The genetic diversity among different gene pools was also estimated in many studies using various marker systems. AFLP marker analyses studies (Iqbal et al., 2001, Abdalla et al.,
Genetic Diversity in Plants

2001, Westengen et al., 2005) revealed that the genetic distance between *G. barbadense* and *G. hirsutum* was in the range of 21-33%. The other wild AD tetraploids (*G. mustelinum, G. tomentosum*) were close to the cultivated AD cottons sharing 75-84% similarity, where *G. tomentosum* was closer to *G. hirsutum* genome (GD=0.16) than the other allotetraploid species (Westengen et al., 2005). At the same time, as mentioned above, *G. darwinii* was closer to *G. barbadense* than *G. hirsutum* (Wendel & Percy, 1990).

Based on AFLP marker analysis, the genetic distance between the widely cultivated AD cottons (*G. barbadense* and *G. hirsutum*) and A-genome diploids varied from 45 to 69%, and that between the cultivated AD cottons and the D-genome varied from 55 to 71%. The genetic distance between the wild AD tetraploids and the A-genome was in the range of 46-52%, and between the wild AD cottons and the D-genome was 58-59%. The genetic distance between the A- and D-genome cottons was in the range of 0.72-0.82 when analyzed with AFLPs (Iqbal et al., 2001, Abdalla et al., 2001, Westengen et al., 2005).

The use of SSR markers revealed that the genetic distance between *G. hirsutum* and *G. barbadense* was in a range of 42-54% (Kebede et al., 2007). However, Lacape et al. (2007) reported higher genome dissimilarity values (D=0.89-0.91%) between *G. hirsutum* and *G. barbadense* within their material. Also, high mean dissimilarity values were reported between *G. hirsutum* and *G. tomentosum* (D=0.71-0.75) and between *G. barbadense* and *G. tomentosum* (D=0.80) using highly polymorphic sets of SSRs (Lacape et al., 2007). The genetic distance among the AD tetraploids was also in the range of 0.80-0.88 (Liu et al., 2000) with moderate closeness of *G. tomentosum* to the Upland cotton than *G. barbadense* cultivars that was also supported by other studies with different marker systems (Dejoode & Wendel, 1992; Hawkins et al., 2005). Based on SSR marker analysis, the genetic distance between the cultivated AD cottons and the A-genome was in the range of 31-43%, and that between the cultivated AD cottons and the D-genome was in the range of 35-46% (Kebede et al., 2007). The genetic distance between A-and D-genome cottons varied in the range of 29-42% (Kebede et al., 2007).

7. Perspectives of 21st century cotton genomics efforts in characterizing and exploiting the genetic diversity of *Gossypium* species

During the past two decades, the international cotton research community has made extensive efforts to utilize the genetic diversity in cotton, which are imperative for the future of trait improvements of the cotton crop. There are many marker systems such as isozymes, RAPDs, RFLPs, AFLPs (extensively referenced herein), and their various modifications (Zhang et al., 2005b) successfully used in cotton. However, the development of a large collection of robust, portable, and PCR-based molecular marker resources such as Simple Sequence Repeats (SSRs; www.cottonmarker.org) and Single Nucleotide Polymorphisms (SNPs) for cotton were one of the tremendous accomplishments of cotton research community (Chen et al., 2007; Van Deynze et al., 2009). This accelerated studies on genetic diversity in cotton at genomic level. Cotton marker resources were made available for cotton research community through cotton marker database (CMD) (Blenda et al., 2006) that are being extensively used to create cotton genetic linkage maps and to map important agronomic QTLs (Abdurakhmonov, 2007; Chen et al., 2007; Zhang et al., 2008). In addition to available DNA marker systems, recently, Reddy et al. (2011) developed a diversity array technology (DArT) marker platform for the cotton genome and evaluated the use of DArT...
markers compared with AFLP markers in mapping populations. These studies are very important to elucidate molecular basis of genetic diversities in cotton that are vital to mobilize useful genes of agronomic importance to the elite cultivars through marker-assisted breeding programs.

Furthermore, researchers have reported several potential candidate genes of many agronomic traits in cotton. Tremendous efforts were made to study molecular basis of one of the most complex, but important traits – cotton fiber development (Abdurakhmonov, 2007; Chen et al., 2007; Zhang et al., 2008). These efforts, including many more recent reports on the dissection of candidate genes that are specifically expressed in developing fibers are undoubtedly imperative for future exploitation of genetic diversity in cotton fiber traits using transgenomics approaches (Arpat et al., 2004; Ruan et al., 2003; Zhang et al., 2011b).

Despite wide spectra of genetic diversity in Gossypium genus and extensive cotton genomics efforts, cotton lags behind other major crops for marker-assisted breeding due to limited polymorphism in the cultivated germplasm. This underlies broadening of cultivar germplasm genetic base through mobilization of useful gene variants from other gene pools into cultivated germplasm. There is a need for application of modern innovative genomics tools such as association mapping to identify genetic causatives of natural variations preserved in cotton germplasm resources and their use in plant breeding. Efforts on turning the gene-tagging efforts from bi-parental crosses to natural population or germplasm collections, and from now classical QTL-mapping approach to modern linkage disequilibrium (LD)-based association study should lead to elucidation of ex situ conserved natural genetic diversity of worldwide cotton germplasm resources and its effective utilization. LD refers to a historically reduced (non-equilibrium) level of the recombination of specific alleles at different loci controlling particular genetic variations in a population (Abdurakhmonov & Abdukarimov, 2008). Although novel to cotton research, the association genetics strategy is, in fact, highly applicable to the identification of markers linked to fiber quality and yield through the examination of linkage disequilibrium (LD) of DNA-based markers with fiber quality and yield traits in a large, diverse germplasm collection (Abdurakhmonov et al., 2004, 2008, 2009).

Application of association mapping strategy in gene mapping and germplasm characterization gained wider use in cotton. For example, Kantartzı & Stewart (2008) conducted association analysis for the main fiber traits in 56 G. arboreum germplasm accessions introduced from nine regions of Africa, Asia and Europe using 98 SSR markers. Association mapping strategy was also applied for tagging fiber traits in the exotic germplasm derived from multiple crosses among Gossypium tetraploid species (Zeng et al., 2009). Both of these studies did not quantify the LD level in the population and used marker-trait associations to tag genetic variations contributing to the trait of interest.

Alternatively, to better assess and exploit a molecular diversity of cotton genus, we conducted molecular genetic analyses in a global set of ~1000 G. hirsutum L. accessions, one of the widely grown allotetraploid cotton species, from Uzbek cotton germplasm collection. This global set represented at least 37 cotton growing countries and 8 breeding ecotypes as well as wild landrace stock accessions. The important fiber quality (fiber length and strength, Micronaire, uniformity, reflectance, elongation, etc.) traits were measured in two distinct environments of Uzbekistan and Mexico. This study allowed us to quantify the linkage disequilibrium level in the genome of Upland cotton germplasm and to design an
“association mapping” study to find biologically meaningful marker-trait associations for important fiber quality traits that accounts for population confounding effects (Yu et al., 2006; Abdurakhmonov & Abdukarimov, 2008). Several SSR markers associated with major fiber quality traits along with donor accessions were identified and selected for MAS programs (Abdurakhmonov et al., 2008, 2009).

Further, with the specific objective of introducing and enriching the currently-applied traditional breeding approaches with more efficient modern MAS tools in Uzbekistan, we began marker-assisted selection efforts based on our association mapping studies mentioned above. For this purpose, we selected (1) a set of twenty three major (Micronaire, fiber strength and length, and elongation) fiber trait-associated DNA markers as a tool to manipulate the transfer of QTL loci during a genetic hybridization; and (2) thirty-seven (11 wild race stocks and 26 variety accessions from diverse ecotypes) donor cotton genotypes that bear important QTLs for fiber traits. These donor genotypes were crossed with 9 commercial cultivars of Uzbekistan (as recipients) in various combinations with the objective of improving one or more of fiber characteristics of these recipients. These 9 parental recipient genomes were first screened with our DNA-marker panel to compare with 37 donor genotypes. The polymorphic status of marker bands between donor and recipient genotypes were recorded. The hybrid plants generated from each crossing combination were tested using DNA-markers at the seedling stage, and hybrids bearing DNA-marker bands from donor plants were selected for further backcross breeding (Abdurakhmonov et al., 2011).

Testing the major fiber quality traits using HVI in trait-associated marker-band-bearing hybrids revealed that mobilization of the specific marker bands from donors really had positively improved the trait of interest in recipient genotypes (data not shown). Currently, we developed a second generation of recurrent parent backcrossed hybrids (F1BC2), bearing novel marker bands and having superior fiber quality compared to original recipient parent (lacking trait-associated SSR bands). These results showed the functionality of the trait-associated SSR markers detected in our association mapping efforts in diverse set of Upland cotton germplasm. Using these effective molecular markers as a breeding tool, we aim to pyramid major fiber quality traits into single genotype of several commercial Upland cotton cultivars of Uzbekistan. Our efforts will not only help rapid introgression of novel polymorphisms, broadening the genetic diversity of cotton cultivars and accelerating the breeding efforts for future sustainable cotton production in Uzbekistan but also exemplify effective exploitation of the natural genetic diversity \textit{ex situ} preserved in cotton germplasm collections (Abdurakhmonov et al., 2011).

In spite of successful application of association mapping in cotton, there is a great challenge with assigning correct allelic relationships (identity by decent) of multiple band amplicons when diverse, reticulated, and polyploid cotton germplasm resources lacking historical pedigree information are investigated. Besides, there is the issue of rare and unique alleles that is problematic for conducting association mapping (Abdurakhmonov & Abdukarimov, 2008). While these issues can be solved using many available methodologies and approaches (Abdurakhmonov & Abdukarimov, 2008); however, recent studies in model crops suggested a new methodology to minimize these issues with the creation of segregating populations, performing genetic crosses between several reference populations with known allele frequencies for functional polymorphisms. Such an approach is referred to as nested
association mapping (NAM) and NAM populations would greatly enhance the power of association mapping in plants (Stich and Melchinger, 2010). The usefulness and feasibility of NAM population based genetic mapping studies were successfully demonstrated in maize (Kump et al., 2011; Poland et al., 2011; Tian et al., 2011;) and should be adopted for other crops with complex genome and diverse germplasm resources like cotton. Therefore, creation of NAM populations for cotton on the basis of germplasm evaluation and characterization studies is the task of high priority for future characterization and mapping of biologically meaningful genetic variations in cotton. This requires further efforts and investments that facilitate fine-scale association mapping studies in cotton. This will ultimately lead to cloning and characterization of genetic causatives controlling the genetic diversities and its effective exploitation in plant breeding.

8. Conclusions

In conclusion, by having a wide geographic and ecological dispersal, the Gossypium genus represents and preserves large amplitude of morphobiological and genetic diversity within its ex situ worldwide germplasm collections and in situ occupation sites. Because of the development of molecular marker technologies, and their application in genetic diversity studies of germplasm resources, various gene pools and specific cultivar groups, researchers found a genetic bottleneck in cultivated cotton germplasm resources. However, there is moderately high molecular diversity present in some specific cultivar germplasm analyzed worldwide, suggesting a need for continual efforts on searching the diverse cultivar germplasm resources using molecular markers. There is a need to extend molecular marker-based diversity studies for tertiary gene pool accessions of cotton. There also is high genetic diversity available within exotic land race stocks, wild AD cottons, and putative A-and D-genome ancestors to AD cottons that have potential to search for genetic variations useful in future improvement of cotton. The variations observed in genetic distance estimations between different studies could be due to (1) germplasm resources chosen for the study, (2) number of accessions analyzed with molecular markers, (3) number of markers and marker types used, (4) genomic regions screened, and (5) subjective features of data analyses process, e.g., considering or removing unique or rare alleles, largely influencing the genetic distance measures.

Further, the narrowness of genetic diversity in cultivar germplasm was associated with recent and possibly future declines in cotton production and its quality, which was a timely warning to accelerate efforts on broadening the genetic base of cultivar germplasm resource via mobilizing novel genetic variants from wild, primitive, pre-domesticated primary, secondary and tertiary gene pools. Traditional efforts have succeeded in introgressing many new genetic variations into cultivar germplasm from other gene pools, but it is still challenging and the breeding process is slow due to a number of genetic barriers and obstacles, as highlighted above, to accomplish the goal. This underlies the importance of development of innovative tools to exploit the biologically meaningful genetic variations, existing in Gossypium genus. The most effective utilization of genetic diversity of Gossypium species further requires characterization of candidate gene(s) underlying the phenotypic and agronomic diversities, acceleration of linkage mapping, map-based cloning and marker-assisted selection that underlie development of modern genomics technologies such as high-resolution, cost effective LD-based association mapping for cotton with its optimization.
through development of modern nested association mapping populations. The development of efficient cotton transgenomics tools and complete sequencing of cotton genome(s) will further accelerate exploitation of genetic diversity in highly specific manner and with clear vision. Future application of whole genome-association strategy with epigenomics perspectives, which currently is widely being applied in human and the other model plants such as Arabidopsis, will have a significant impact on identifying true functions of genes controlling available genetic diversity, and consequently, its effective utilization.

9. Acknowledgements

We thank the Office of International Research Programs (OIRP) of United States Department of Agriculture (USDA) for continual funding of our collaborative research on cotton germplasm characterization and genetic diversity analysis. We acknowledge Civilian Research Development Foundation (CRDF), USA for project coordination and Academy of Sciences of Uzbekistan for their continual in-house support of the research efforts.

10. References


www.intechopen.com
Genetic Diversity in Gossypium genus

333


Cotton (Gossypium hirsutum) from Gossypium longicalyx. Crop Science, Vol. 47, No. 5, (September-October 2007), pp. 1865-1877, ISSN 0011-183X


Genetic Diversity in Gossypium genus


Genetic diversity is of fundamental importance in the continuity of a species as it provides the necessary adaptation to the prevailing biotic and abiotic environmental conditions, and enables change in the genetic composition to cope with changes in the environment. Genetic Diversity in Plants presents chapters revealing the magnitude of genetic variation existing in plant populations. The increasing availability of PCR-based molecular markers allows the detailed analyses and evaluation of genetic diversity in plants and also, the detection of genes influencing economically important traits. The purpose of the book is to provide a glimpse into the dynamic process of genetic variation by presenting the thoughts of scientists who are engaged in the generation of new ideas and techniques employed for the assessment of genetic diversity, often from very different perspectives. The book should prove useful to students, researchers, and experts in the area of conservation biology, genetic diversity, and molecular biology.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:
