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

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

BOOK CITATION INDEX
CLARIVATE ANALYTICS
INDEXED

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

The Diploid D Genome Cottons (*Gossypium* spp.) of the New World

Mauricio Ulloa

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/58387

1. Introduction

The genetic diversity of cotton (*Gossypium* spp.) is exclusively wide with diverse geographic and ecological niches [1]. The *Gossypium* genus belongs to the *Malvaceae* family. This genus contains more than 45 diploid species and five well-documented allotetraploid species [2-4]. Species of this genus are grouped into nine genomic types ($x=n$=13, $2n$=26 diploid, and $4x$=52 tetraploid) with the following designations: AD, A, B, C, D, E, F, G, and K [3]. Genomic designations are based on the similarities in chromosome size and structure, and the success of interspecific crosses. Based on their chromosomal uniformity, the diploid D genome species of the New World include 26 somatic chromosomes. Some hybrids within genomes are fertile and their chromosomes recombine during meiosis. However, hybrids across genomes are generally infertile and they have a few stable bivalents at meiosis as a result progeny-plant survival from interspecific crosses is sometime low [4]. The allotetraploid cottons [Upland, *G. hirsutum* (AD$_1$) and *G. barbadense* (AD$_2$)] of the New World dominate world natural-fiber production. And they can be described as large shrubs to trees [3,5]. An allotetraploid is a species that derived from the combination of two different genomes or doubling of genomes that are different. The At subgenome is probably best represented by a composite of two diploid genomes ([*G. herbaceum* L. (A$_1$) and *G. arboreum* L. (A$_2$)] from the Old World. These Asiatic species-progenitor cottons primarily produce fibers for non-industrial-textile consumption in India and Asia [4]. The Dt subgenome has a more complex genome (D) of the diploid species-progenitors from the New World. The D genome is comprised of formally reported 13 species [3,7-9] and several undescribed taxa e.g. US-72 [8-11]. Eleven of the 13 species of the New World reside in the country of Mexico (Fig. 1). Taxonomically, these species are recognized as the *Houzingenia* subgenus [5,7]. None of these D genome diploid species produce commercial fibers. These species of the D genome are not well known to public and

private breeding programs around the world, and their utilization for cotton improvement has not been fully exploited. Some species of the D genome [*G. aridum* (D_4), *G. lobatum* (D_7), *G. laxum* (D_9), etc] with arborescent growth habits express unique flowering and fruiting habits (following defoliation in the dry season). And even though none of the D diploid species produce commercial fibers, the diploid D genome species of the New World harbor important genes for improving fiber quality, pest and disease resistance, and drought and salt tolerance in the modern cultivated Upland and Pima cottons.
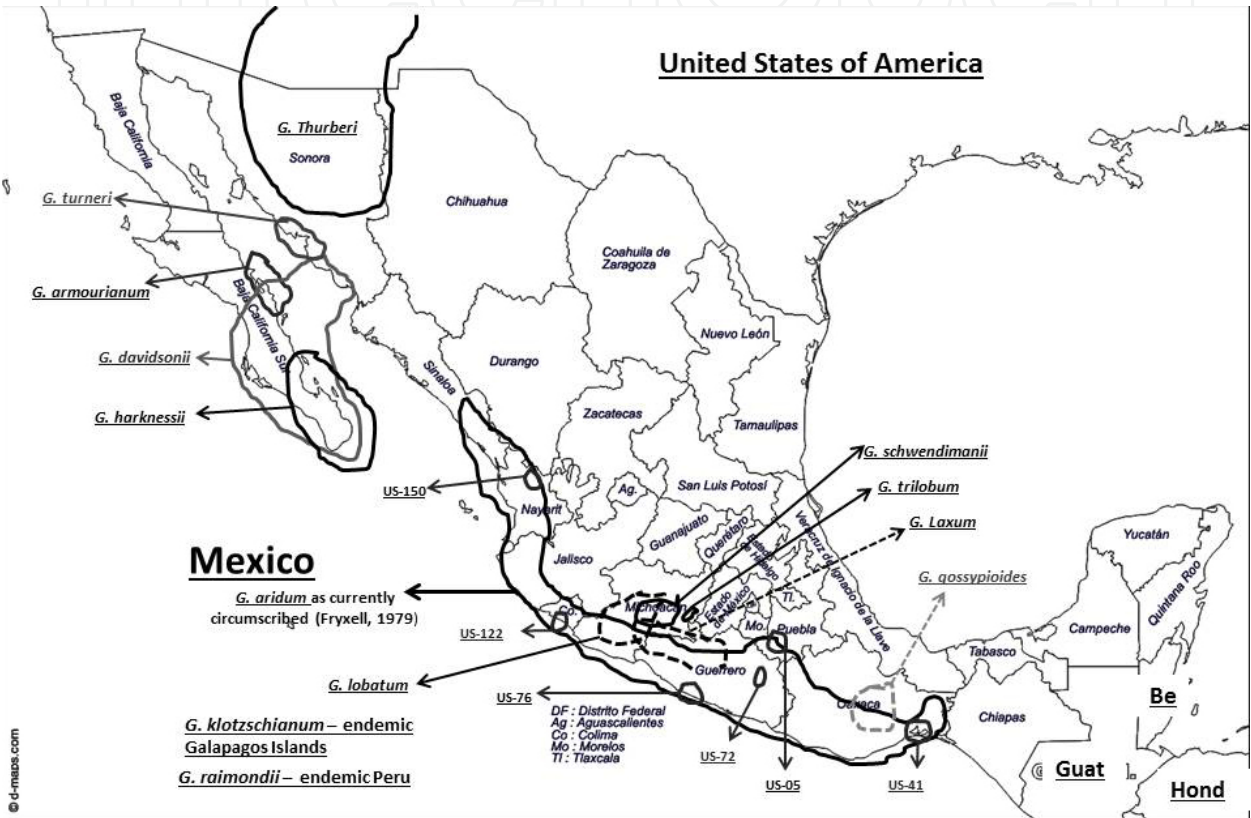
Figure 1. The country of Mexico, states, and approximate areas-boundaries of the diploid species of the D genome endemic to this country. The location of new taxa with US-XX (XX=number – Table 2) is indicated representing possible new species.

Even though Mexico's natural heritage of cotton genetic resources equals that of maize, until recently no national resources were dedicated to the preservation of this natural treasure [8]. The collection/exploration trips of these species have been difficult to document. Increasing human population and urbanization have severely reduced the survival of some of these species. *In situ* conservation of some of these species is threatened. New roads and population growth continue to increase. At this point, one species (*G. aridum* as formally reported) of the subsection *Erioxylum* appears not to be threatened, probably because of the great diversity (botanical and geographic) encompassed by this species. However, some of the most recent collected and non-described taxons (e.g., US-72) or ecotypes of the *G. aridum* species may be in the process of becoming extinct in the wild. In addition, the D_8 *G. trilobum* species is almost

extinct or already extinct. If *in situ* diversity of the Mexican cottons is severely eroded, the germplasm collections all over the world and the USDA Cotton Germplasm Collection will assume a highly significant role in the preservation of the diversity previously residing in Mexico's cotton species of the D genome.

Recently, the genome sequence of the best model, closest living ancestor relative of the allotetraploid cottons-of the Dt subgenome *(G. raimondii)*, was published. This new information compiled with the ongoing next generation sequencing (NGS) projects around the world will provide insights into the evolution, population structure, genetic diversity, and utilization of this genetic resource. The next generation of genomic research will sequence characterize and locate genes that will help molecular breeders to identify differences among germplasm and breeding lines and to apply traditional genetic analyses to infer genes for marker assisted selection (MAS). In addition, the new sequence information obtained through NGS will be an important resource to improve the cotton crop through transgenic technology.

2. Classification of the D diploid species, distribution, and dissemination

The country of Mexico, besides being a part of the center of origin/diversity of *G. hirsutum,* also harbored 11 out of the 13 formally reported D species [3,7] and several non-described taxa of the New World diploid *Gossypium* species (one non-described taxon US-72,) [8-10]. The species of the *Houzingenia* subgenus are classified into six subsections: subsection *Houzingenia* Fryxell [*G. thurberi* Todaro (D_1) and *G. trilobum* (Mociño & Sessé ex DC.) Skovsted (D_8)]; subsection *Integrifolia* Todaro [*G. davidsonii* Kellogg (D_{3-d}) and *G. klotzschianum* Andersson (D_{3-k})]; subsection *Caducibracteolata* Mauer [*G. armourianum* Kearney (D_{2-1}), *G. harknessii* Brandegee (D_{2-2}), and *G. turneri* Fryxell (D_{10})]; subsection *Erioxylum* Rose & Standley [*G. aridum* (D_4), *G. lobatum* (D_7), *G. laxum* (D_9), and *G. schwendimanii* Fryx. & Koch (D_{11})]; subsection *Selera* (Ulbrich) Fryxell [*G. gossypioides* (D_6)], and subsection *Austroamericana* Fryxell [*G. raimondii* Ulbrich (D_5)] [3,4,7]. Eleven of the 13 species of the subgenus *Houzingenia* are distributed in Mexico and extend northward into Arizona (Fig. 1). The other two D species have disjointed distributions; *G. raimondii* is endemic to Peru, while *G. klotzschianum* is found in the Galápagos Islands. The species of the D genome are not well known and utilized in public and private breeding programs around the world. Additional information about morphological characteristics and distribution of the species can be found in Fryxell monograph [12] and several other publications [8-9]. A supplemental information about recent collections [8-9] can be found at the USDA-ARS, SPA, CSRL, Plant Stress and Germplasm Development website (http://www.lbk.ars.usda.gov/psgd/index-cotton.aspx). Also, the Mexican Instituto Nacional de Investigaciones Forestales Agricolas y Pecuarias (INIFAP), Iguala Gro. Mex. nursery has provided us with the opportunity to further study some of these species *ex situ*. Table 1 provides information on 12 of the species during their *ex situ* preservation at the Iguala nursery in Mexico. Data from *G. klotzschianum* is missing from this table. Species planted at the nursery flower from September to January, while *in situ*, some of the populations from these species flower through March-April..

Species	Genome	Petal color	Filament color	Anther color	Leaf shape	Number of seed per capsule	Seed Size length mm	Seed width mm
G. aridum	D_4	Cherry	Purple	Purple	Cordate	14	7.3	1.7
G. armourianum	D_2	Yellow	White	Cherry	Ovate	13	6.0	1.7
G. davidsonii	D_{3-d}	Yellow	Yellow	Yellow	Cordate	14	4.8	1.8
G. gossypioides	D_6	Cherry	Purple	Purple	Palmate	9	5.0	1.8
G. harknessii	D_{2-2}	Yellow	Amarlla	Yellow	Cordate	7	6.6	4.3
G. laxum	D_9	Cherry	Purple	Purple	Lobed	6	8.8	1.6
G. lobatum	D_7	Cherry	Purple	Purple	Lobed	8	10	1.8
G. schwendimanii	D_{11}	Cherry	Purple	Purple	Cordate	7	10	2.5
G. raimondii	D_5	Yellow	Purple	Purple	Cordate	12	7.0	3.6
G. thurberi	D_1	Yellow	White	Yellow	Palmate	14	3.3	1.6
G. trilobum	D_8	Yellow	Yellow	Yellow	Lobed - Palmate	12	4.0	1.2
G. turneri	D_{10}	Yellow	Yellow	Yellow	Cordate	10	3.8	1.8

Table 1. Averages of flowers and seed characteristics taken on some of the *Gossypium* species of the D genome established in the permanent nursery at Iguala, Guerrero Mexico.

*Gossypium aridum***(D_4),** as formally reported, is the most widely distributed wild *Gossypium* in Mexico [5,9]. The distribution of this species, non-described taxa (one non-described taxon US-72, [8-11]), and described ecotypes [8-9,11] of the New World extends from the northern state of Sinaloa to the southern state of Oaxaca (Fig. 1). The species and taxa/ecotypes can be described as medium to large trees from five to 18 m tall or larger. As expected from its wide range, this *Gossypium* species occupies a number of habitat niches (http://www.lbk.ars.usda.gov/psgd/index-cotton.aspx). Comparisons among specimens of on-site observations indicate extensive differences in leaf size, vestiture of the leaves, morphology in the lysigenous glands on the capsules, and period of flowering. Morphologically, the leaves of the ecotype from Oaxaca are the largest of the species, with a relatively dense but fine indumentum. Several populations of *G. aridum* that appear to be very similar in morphology are distributed along the coastal foothills of Jalisco, Colima, Guerrero, and possibly Michoacán. Figure 2 presents an accessions of a population collected from recent exploration/collection trips [8-9]. The elevation of these populations range from <60 m up to >1000 m. Generally, these populations have almost no leaf trichomes, and small mature capsules (Table 1). Also, flowering in the states of Sinaloa, Nayarit and Jalisco is delayed until March and April in these types, and their capsules do not mature until late April-May (Fig. 2). While the Coastal

populations in the states of Colima, Guerrero, and Oaxaca mature their capsules in February and March. Well-documented populations of collected *G. aridum* have now been made from several regions representing different non-described taxa and ecotypes [9]. These collections will continue to allow for *ex situ* preservation, maintenance and evaluation. This will also allow common-garden comparisons of all populations.

Figure 2. Habitat, trees, flower, capsule, and leaves of collected accessions of the *G. aridum* species.

For the most part, populations of *G. aridum* occur as a part of the native vegetation in deciduous woodlands. Different non-described taxa and ecotypes appear to thrive in areas where the woodland is disturbed, particularly along road banks where the canopy is opened. In the niches where they occur, some of these populations are usually found in abundance, although these locations may be separated by many kilometers. This species, as presently circumscribed, is very diverse (non-described taxa/ecotypes) and some of these populations do not appear to be threatened (Fig. 1).

*Gossypium armourianum***(D$_2$)** is distributed from Baja California to the Gulf of California on the San Marcos Island (Fig. 1). This species can be described as a compact branched shrub of around one m tall. The species for the most part contains ovate leavesyellow flowers with withe

filaments and cherry anthers. The seeds are contained in a capsule with three to four cells, and seeds averaging 6.0 x 1.7 mm with brownish tightly compressed fibers (Table 1; [12]).

Figure 3. A two-year shrub of the species of *G. davidsonii*.

Gossypium davidsonii(D_{3-d}) is adapted to the desert environments of the southern Baja California peninsula and across the Gulf of California in the state of Sonora. This species is described as a branched shrub of one to two m tall and for the most part with cordate leaves. Figure 3 presents one of the accessions maintained at the Iguala nursery with cordate leaves. This species has flowers with yellow colored petals, filaments, and anthers. The seeds are contained in a capsule commonly with four cells and seeds averaging 4.8 x 1.8 mm with sparse compressed fibers (Table 1; [12]).

Gossypium gossypioides(D_6) is distributed in the central part of the state of Oaxaca and is adapted to a higher altitude than any other arborescent D *Gossypium* species, >1000 m. This species has been encountered only in the state of Oaxaca. It has been hypothesized that the distribution of *G. gossypioides* may be strongly influenced by elevation. One aspect of *G. gossypioides* that was recently reported is its deciduous habit as a drought escaping mechanism [8]. This species, like the other arborescent *Gossypium* species in the section *Erioxylum*, occurs in dry deciduous woodlands of Oaxaca and defoliates with the onset of the dry season. However, unlike the species of subsection *Erioxylum*, it flowers and fruits near the end of the wet season before defoliating. Fryxell [5] was unaware of the deciduous nature of the foliage similar to the other arborescent species of Mexico, which defoliated as a mechanism to escape drought. This species is comprised of small trees from three to seven m tall. Figure 4 presents one of the accessions (one year old) maintained at the Iguala nursery with flowers with light cherry color, and purple filaments and anthers. Also,

this species as *G. raimondii* possess a unique petal mutation called reverse petal spot (pigment is present on adaxial and abaxial petal surfaces). *G. gossypioides* has been reported with cryptic repeated genomic recombination during speciation, with conflicting morphological, cytogenetic, and molecular evidence of its phylogenetic affinity to other New World cottons [13]. Figure 4 presents trees and mature capsules encountered at the natural habitat of this species. The seeds are contained in a capsule with three cells and seeds averaging 5.0 x 1.8 mm with grayish sparse compressed fibers (Table 1; [12]).

Figure 4. One year accession of the species *G. gossypioides* which was maintained at the Iguala Guerrero Mex. nursery with a flower, also, trees and mature capsules encountered at its natural habitat.

Gossypium harknessii(D_{2-2}) is adapted to the desert environments of Baja California in the Cape region and adjacent islands. This species is described as a shrub of around three m tall and for the most part with cordate leaves. Plants present flowers with yellow colored petals, filaments, and anthers. The seeds are contained in a capsule commonly with three-to four cells and seeds averaging 6.6 x 4.3 mm with grayish sparse compressed fibers (Table 1; [12]).

Gossypium klotzschianum(D_{3-k}) is one of the two species with disjointed distribution (not found in the country of Mexico) and is endemic to the Galápagos Islands. This species is described as a shrub up to four m tall and for the most part with petiolate-cordate leaves. Plants maintained at the Iguala nursery present flowers with yellow petals, filaments, and anthers.

The seeds are contained in a capsule commonly with four cells, ciliate on inner suture margins, and seeds averaging 5.0 x 2.5 mm with sparse inconspicuous fibers [12].

*Gossypium laxum***(D₉)** is reported to be located in the Cañon del Zopilote in the central state of Guerrero, and, more recently, it was reported to be found in the state of Michoacan along the road between Huetamo and Nuevo Churumuco. The full range of *G. laxum* is yet to be determined, but it probably extends many kilometers along the Rio Balsas watershed east and west of Cañon del Zopilote. This taxon does well in areas of open sunlight, such as road cuts, like other members of subsection *Erioxylum*. However, it can also be found as part of natural deciduous woodland vegetation. Morphological diversity is not extensive among the accessions that have been collected so far. The collected accession (US-98, D_{9-6-M}, [9]) significantly extended the range of this species to the west. It is adapted to altitudes ranging from 200 to 900 m. Because some habitat sites of this species have generally been found not suitable for agriculture, *G. laxum* may be able to survive Mexico's demographic changes and does not seem to be threatened at present. This species is another arborescent *Gossypium* species of the subsection *Erioxylum* and comprises trees up to 10 m tall *in situ*. It flowers and produces fruits near the end of the wet season before defoliating. Plants of these populations present flowers with light cherry color, and purple filaments and anthers. US-98 accession presented flowers similar to *G. lobatum*. The seeds are contained in a capsule with three-to five-cells and seeds averaging 8.8 x 1.6 mm with densely pubescent fibers (Table 1; [12]).

Gossypium lobatum (D₇) is adapted to the environment of the central state of Michoacan. The collected accession (US-112, D_{7-10-M}, [9]) extended the range of this species to the west of the state. The distribution of this taxon is probably throughout the Rio Tepalcatepec watershed with an eastern extension along the Rio Balsas watershed. *G. lobatum* has unique features such as its distichous leaf insertion and tomentose calyces with prominent lobes, the characteristic from which its name is derived. While all accessions had distichous leaves, the calyces of the western accession-populations were less hairy and the lobes less prominent. It is unknown how much suitable habitat has already been destroyed. Overall the species does not appear to be threatened at present. It is adapted to altitude ranging from 200 to 400 m. This species is another arborescent *Gossypium* species of the subsection *Erioxylum* and comprises trees up to 10 m tall *in situ*. It flowers and fruits near the end of the wet season before defoliating. Plants of these populations present flowers with light cherry color, and purple filaments and anthers. The seeds are contained in a capsule with three-cells and seeds averaging 10.0 x 1.8 mm with densely pubescent fibers (Table 1; [2]).

*Gossypium raimondii***(D₅)** is the other of the two D species that has disjointed distribution and is endemic to Peru. This species was originally described as a shrub two to three m tall. However, accessions planted at the Iguala nursery from seed provided by the USDA-ARS Cotton Germplasm Collection, College Station TX were able to produce shrub like trees up to 10 m tall. For the most part, accessions at the Iguala nursery present large cordate leaves, flowers with yellow petals, and purple filaments and anthers. The seeds are contained in a capsule commonly with four-cells, narrowly ovoid, and seeds averaging 7.0 x 3.6 mm with densely pubescent fibers [12]. *G. raimondii* is considered the closest living ancestor relative of the allotetraploid cottons (Dt subgenome) [4,12].

Gossypium schwendimanii (D$_{11}$) is the most recently described *Gossypium* species of the D genome from Mexico [3,12]. This species was encountered for the first time at the Guerrero-Michoacán border near Infiernillo where a population of *G. aridum* was located only 2 km from "typical" *G. schwendimanii*. Collected accessions from this location showed some morphological features similar to *G. aridum* that suggested some introgression between the two species. Recently, *G. schwendimanii* [9] was collected from an area about 20 km south of *G. lobatum*. Seeds were collected from two additional populations of *G. schwendimanii* from the hills above the west side of Presa Infiernillo. Little morphological diversity was evident among the populations. The full natural distribution or native range of *G. schwendimanii* is unclear because of the limited information available on the species. One factor that has an impact on its genetic identity is its apparent sympatry with *G. aridum* and *G. lobatum* over parts of its range. It is adapted to altitude ranging from 200 to 400 m. This species is another arborescent *Gossypium* species of the subsection *Erioxylum* and comprises trees up to 10 m tall *in situ*. It flowers and produces fruits near the end of the wet season before defoliating. Plants of these populations present flowers with light cherry color, and purple filaments and anthers. The seeds are contained in a capsule with three-or four-cells and seed averaging 10.0 x 2.5 mm with densely pubescent fibers (Table 1; [9]).

Gossypium thurberi (D$_1$) is distributed from the state of Arizona, U.S.A to the state of Sonora, Mexico. This species can be described as a small tree or shrub of around 3 m tall. The species for the most part contains palmate leaves with flowers from white to yellow color, white filaments, and yellow anthers. The seeds are contained in a capsule with three-cells and seeds averaging 3.3 x 1.6 mm with blackish color and no fibers (Table 1; [12]).

Gossypium trilobum (D$_8$) is generally limited to moderately high elevations (1200 m) in western Mexico. This species belongs to section *Houzingenia* and is a sister species to *G. thurberi*, the most northerly (Sonora-Mex to Arizona-U.S.) distributed species of the *Gossypium*. Even though Fryxell [5] indicated that it was widely distributed from the state of Sinoloa to the state of Morelos, few if any collections have been made in the last 30 years in any of these locations. Based on the results of a survey made by Ulloa et al [8-9] special mention should be made of *G. trilobum*. In 2002-2004, sites were visited where herbarium specimens were collected in the past. In each of five widely separated locations represent-ed by herbarium (MEXU) collections (sites in N and W México, Jalisco, Michoacán, and Morelos), *G. trilobum* plants or populations were unable to be located. Although the status of *G. trilobum* in remote areas is unknown, the results from the survey by Ulloa et al [8-9] confirmed that the distribution of this species has been severely eroded by agricultural and human-population pressures. For the most part, the habitat of this species has been replaced by intense and extensive agricultural production of guava (*Psidium* spp.) in the State of Michoacan, Mexico. At this point, *G. trilobum* species is almost extinct or is becoming extinct in the wild. This species can be described as a small tree or shrub of around 4 m tall. The species for the most part contains lobed-palmate leaves, and yellow flowers, filaments and anthers. The seeds are contained in a capsule with three-cells and seeds averaging 4.0 x 1.2 mm with blackish color and compressed pubescence (Table 1; [12]).

Figure 5. Shrubs of *G. turneri* in their natural habitat showing green vegetation after extended drought.

*Gossypium turneri***(D$_{10}$)** is adapted to the coast of the state of Sonora Mpio. of Guaymas and primarily associated with soils of weathered igneous concentration (Fig. 1). On a long-term basis, *G. turneri* is well adapted to the sea-shore environments and sea level altitude in which it occurs. Based on a recent exploration (J.M. Stewart and M. Ulloa, 2004 expedition), the species has salt and drought resistance mechanisms that allow it to survive extended periods without rain. During this trip, the first *G. turneri* bush encountered was actually in the yard/sea-cliff. Although the species had a very small yield of seed the year when it was encountered because of drought, no evidence was seen of plants that died from the lack of water. Like most other *Gossypium*, in those areas where the species occurs, the plants are quite numerous. Unless unforeseen expansion of the resort industry occurs along the coastal region north of La Manga, the species habitat, in general, probably will not undergo rapid degradation. If resort con-struction should occur on the sea cliffs and adjoining valleys, then the species most likely would be lost in the wild. This species can be described as a shrub of around 1 m tall. Figure 5 presents shrubs in their natural habitat still showing green vegetation after extended drought. The species for the most part contains cordate leaves with yellow flowers, filaments and anthers. The seeds are contained in a capsule with three-to five-cells and seeds averaging 3.8 x 1.8 mm with blackish color and compressed pubescence (Table 1; [12]).

3. Collections/explorations and unclassified taxa

The gene pool of Upland/Acala *G. hirsutum* from the country of Mexico derived one of the primary sources for improvement of most of the Acala and Upland cotton growing in the world today. In addition, another cotton genetic resource of this country is the 11 formally reported D diploid *Gossypium* species and several unclassified taxa [8-11] of the Western Hemisphere. Mexico and its boundaries are the center of diversity of these endemic species. Some of these species and their genomes (US-72, D_4, D_7, D_9, D_{10}, and D_{11}) with arborescent or shrub growth habits express unique flowering and fruiting habits (following defoliation in the dry season) and salt and drought resistance mechanisms that allow them to survive extended periods without rain.

Because of the importance of the gene pool of *G. hirsutum* from Mexico, the collection/ exploration trips of the D diploid species have been difficult to execute and document. Two of the greatest explorers and taxonomists of the *Gossypium* genus, and especially species from the country of Mexico, were Drs. Fryxell and Stewart. P.A. Fryxell made several collection-expeditions from 1968 to 1975 in the country, providing a larger number of specimens to several Herbariums (Herbarium Nacional de Mexico-MEXU and Herbarium Instituto de Ecologia A.C. Mexico-XAL) with clear and precise descriptions of habitat and location of collected accessions [5]. He also made the most recent taxonomic classification of *Gossypium* species [7]. A. E. Percival, J. M. Stewart (USDA), A. Hernandez and F. de Leon (INIFAP) made several collection-expeditions in 1984 throughout the states of the Yucatan Peninsula and in parts of the states of Tamaulipas, Veracruz, Tabasco, Oaxaca and Chiapas. Also, A. E. Percival (USDA-ARS), J. M. Stewart (Univ. of Arkansas), E.A. Garcia, and L. Peréz (INIFAP, Mexico) made additional collection-expeditions in 1990 in the state of Baja California Sur and parts of the states of Sonora and Sinaloa. As a result of their early efforts, a number of *Gossypium* accessions of the subgenus *Houzingenia* from various parts of Mexico were deposited in the USDA-ARS Cotton Germplasm Collection College Station, TX, USA. Also, during the 1980s, Dr. Lemeshev of the Academy of Science of Russia established a *Gossypium* nursery in Iguala City, state of Guerrero in the country of Mexico. Also, some or all of these species are catalogued in the germplasm collection of the Vavilov Institute in St. Petersburg and in several collections of Former Soviet Union countries (e.g., Uzbekistan) based on several collection-expeditions by the Universidad Autónoma de Guerrero Mexico and the Academy of Science of Russia in the states of Veracruz, Tabasco, Campeche, Yucatán, Chiapas, Guerrero, Oaxaca, Michoacán, Morelos, Colima, Sinaloa, Sonora and Baja California Sur between 1989 and 1993 by F. Talipov, C. Cataláio, F. Salgado and M. Bahena. This nursery was abandoned upon Dr. Lemeshev's return to Russia (Q. Obispo, personal communication; [8]).

Until recently no national resources were dedicated to the preservation of this natural treasure [8-9]. In 2002-2006, the United States Department of Agriculture – Agriculture Research Service (USDA-ARS) and the Mexican Instituto Nacional de Investigaciónes Forestales Agricolas y Pecuarias (INIFAP) sponsored joint *Gossypium* germplasm collection trips by U.S. and Mexican cotton scientists. As a result of these efforts, a significant number of *Gossypium* accessions of the subgenus *Houzingenia* from various parts of Mexico were collected (Table 2). Collected

No. of Accessions	Species	Genome	Entry ID
2	G. hirsutum	AD_1	TM-1 and Acala Maxxa
1	G. barbadense	AD_2	Pima 3-79
1	G. tomentosum	AD_3	G-tom
1	G. mustelinum	AD_4	G-must
1	G. darwinii	AD_5	G-darw
12	G. herbaceum	A_1	A_{1-8-1}, A_{1-8-2}, A_{1-5}, A_{1-9}, A_{1-17}, A_{1-18}, A_{1-19}, A_{1-22}, A_{1-23}, A_{1-40}, A_{1-49}, and A_{1-52}
11	G. arboreum	A_2	A_{2-8}, A_{2-41}, A_{2-47}, A_{2-61}, A_{2-72}, A_{2-82}, A_{2-106}, A_{2-141}, A_{2-194}, A_{2-234}, and A_{2-241}
7	G. thurberi	D_1	D_1, D_{1-4}, D_{1-23}, D_{1-24}, D_{1-35}, D_{1-37}, and $D_{1-35XD8-6}$
5	G. armourianum	D_2	D_{2-1}, D_{2-2}, D_{2-q}, D_{2-w}, and $D_{2-19XD2-17}$
5	G. davidsonii	D_3	D_{3-1}, D_{3-2}, D_{3-23}, D_{3-26}, and D_{3-28}
32	G. aridum	D_4	D_{4-1-P} (US004), D_{4-2-P} (US005), D_{4-3-O} (US010), D_{4-4-O} (US011), D_{4-5-O} (US012), D_{4-6-O} (US013), D_{4-7-O} (US15), D_{4-8-O} (US016), D_{4-9-O} (US017), D_{4-10-O} (US041), D_{4-11-G} (US072), D_{4-12-G} (US076), D_{4-13-G} (US078), D_{4-14-G} (US080), D_{4-15-G} (US081), D_{4-16-C} (US117), D_{4-17-C} (US120), D_{4-18-C} (US121), D_{4-19-C} (US122), D_{4-20-C} (US126), D_{4-21-J} (US128), D_{4-22-J} (US130), D_{4-23-J} (US136), D_{4-24-N} (US138), D_{4-25-C} (D4-168a), D_{4-26-C} (D4-168b), D_{4-27-C} (D4-168c), D_{4-28-N} (US147), D_{4-29-N} (US148-a), D_{4-30-N} (US148-b), D_{4-31-N} (US149), and D_{4-32-N} (US150)
3	G. raimondii	D_5	D_{5-1}, D_{5-2}, and D_{5-3}
2	G. gossypioides	D_6	D_{6-1-O} (US043) and D_{6-2-O} (US046)
10	G. lobatum	D_7	D_{7-1-M} (US086), D_{7-2-M} (US101), D_{7-3-M} (US103), D_{7-4-M} (US104), D_{7-5-M} (US105), D_{7-6-M} (US106), D_{7-7-M} (US109), D_{7-8-M} (US110), D_{7-9-M} (US111), and D_{7-10-M} (US112)
9	G. trilobum	D_8	D_{8-1-M} (US160), D_{8-2-M} (US162), D_{8-3-M} (US163), D_{8-A}, D_{8-B}, D_{8-1}, D_{8-6}, D_{8-10}, and $D_{8-6XD1-35}$
5	G. laxum	D_9	D_{9-1-G} (US065), D_{9-2-G} (US066), D_{9-4-G} (US068), D_{9-5-G} (US070), and D_{9-6-M} (US098)
1	G. turneri	D_{10}	D_{10-1-S} (US156)
3	G. shwendimanii	D_{11}	D_{11-1-M} (US083), D_{11-2M} (US084), and D_{11-3M} (US100)

[a]Letter at the end of the Entry ID indicating the state where the accession was collected: C=colima, G=Guerrero, J=Jalisco, N=Nayarit, M=Michoacan, O=Oaxaca, P = Puebla , and S = Sonora ecotypes. This table was published in the following article: Ulloa M, Abdurakhmonov IY, Perez-M C, Percy R, Stewart McDJ. Genetic diversity and population structure of cotton (Gossypium spp.) of the New World assessed by SSR markers. Botany 2013;91 251-259. License agreement was provided by the NRC Research Press – Copyright Clearance Center.

Table 2. Summary of the 111 accessions used to investigate genetic diversity and population structure of the New World cottons (*Gossypium* spp.).

accessions were placed in a nursery or botanical garden in Iguala, Guerrero, Mexico, including several accessions of each of the arborescent species for ex *situ* conservation. Today, Mexico maintains this *Gossypium* nursery in Iguala, Guerrero (C. Perez-Mendoza. *personal communication*). Since the first collection-expedition trips were made, the *in situ* survival of these diploid species has been threatened by increasing human population, modernization of agriculture and urbanization. If *in situ* diversity of the Mexican cottons is severely eroded, then current and additional accessions in all the germplasm collections all over the world and the USDA Cotton Germplasm Collection assume a highly significant role in the preservation of the diversity previously residing in Mexico's dooryard (*G. hirsutum*) and cotton species of the D genome.

As formally reported, *G. aridum* is the most widely distributed wild *Gossypium* species in Mexico [5,7]. However, recent additional collections (http://www.lbk.ars.usda.gov/psgd/index-cotton.aspx) and several studies [8-11] have reported and suggested non-described taxa and ecotypes that can be considered separate species. Morphological comparisons among specimens of on-site observations indicate extensive differences in leaf size, vestiture of the leaves, morphology in the lysigenous glands on the capsules, and period of flowering. Molecular comparisons have provided useful information in the geographical, taxonomical distribution, and evolutionary history of the New World D genome-species. Based on molecular data (see additional information in the section below), in addition to previously D_{4-11-G}, US-072 reported new taxon, five new collected accessions [D_{4-10-O} (US-41), D_{4-2-P} (US-05), D_{4-12-G} (US-76), D_{4-19-C} (US-122), and D_{4-32-N} (US-150)] from five different geographical sites (ecotypes) from the states of Oaxaca, Puebla, Guerrero, Colima, and Nayarit may be recognized as new species [8-9]. Subsequent observations on greenhouse plants and a return visit to the sites when the plants are beginning to flower will taxonomically confirm that these populations represent undescribed taxa (new species) belonging to subsection *Erioxylum*.

4. Molecular characterization of the D genomes

When traditional taxonomy based on morphology (plant canopy, plant height, leaf and capsule shapes, flowers and petal spots, seed size, etc) do not distinguish two species with intermediate phenotypes, molecular methods provide an alternative solution of resolving these not well defined morphological differences between two species [9,14-15]. Molecular markers such as amplified fragment-length polymorphism (AFLP) [10-11] and microsatellites or simple sequence repeats (SSR) have been used to reveal genetic diversity and to distinguish not well defined differences between species or wild relatives [9]. Molecular marker-gene methods have also been used, for example: internal transcribed spacer (ITS) of ribosomal DNA or ribosomal DNA fragment gene-comparisons [16-18], and fragments of chloroplast DNA [19], or repetitive DNA [20-22]. In addition, a few loci of the *Adh* gene [19,23-25], *FAD*2-1 gene [26], and *Ces A1* gene [27] provided insight into the characterization of the *Gossypium* species. Moreover, the phylogeny of the New World diploid *Gossypium* was analyzed based on three independent single-copy genes (*A1341*, *AdhC*, and *CesA1b*) [28]. These genes were used in previous studies [22-23], showing a high ratio of phylogenetical informative fragment data.

Even though phylogenetic relationships with these three single-copy genes among species of the D genome still remain unclear, the molecular data supported the recognition of a new D species (US-72) closely related to *G. laxum* [28]. Similar observations were obtained when molecular diversity and phylogenetic relationships were examined among 33 accessions of arborescent *Gossypium* including 23 of *G. aridum* with Random Amplified Polymorphic DNA (RAPD) and AFLP fragments [11].

In 2013 Ulloa et al [9] reported a study of genetic diversity and population structure of cottons (*Gossypium* spp.) of the New World (Table 2). In this study, the genetic diversity and population structure of 111 cotton accessions of *Gossypium* were assessed with SSR markers with wide genome coverage. The species represented five allotetraploids ($AD_1 – AD_5$ genomes), 23 Asiatic diploids of the Old World (A_1 and A_2 genomes), and 82 diploids of the New World subgenus *Houzingenia* ($D_1 – D_{11}$ genomes) species (Table 2).
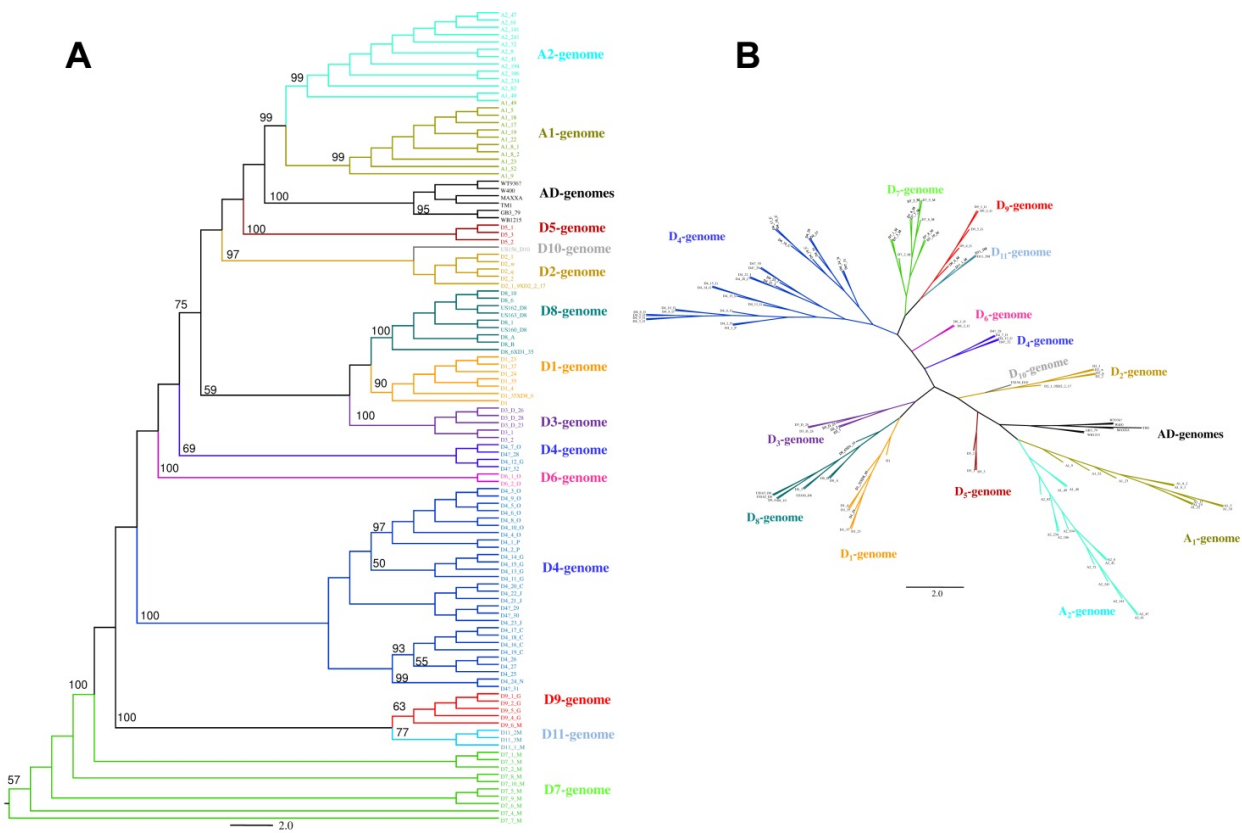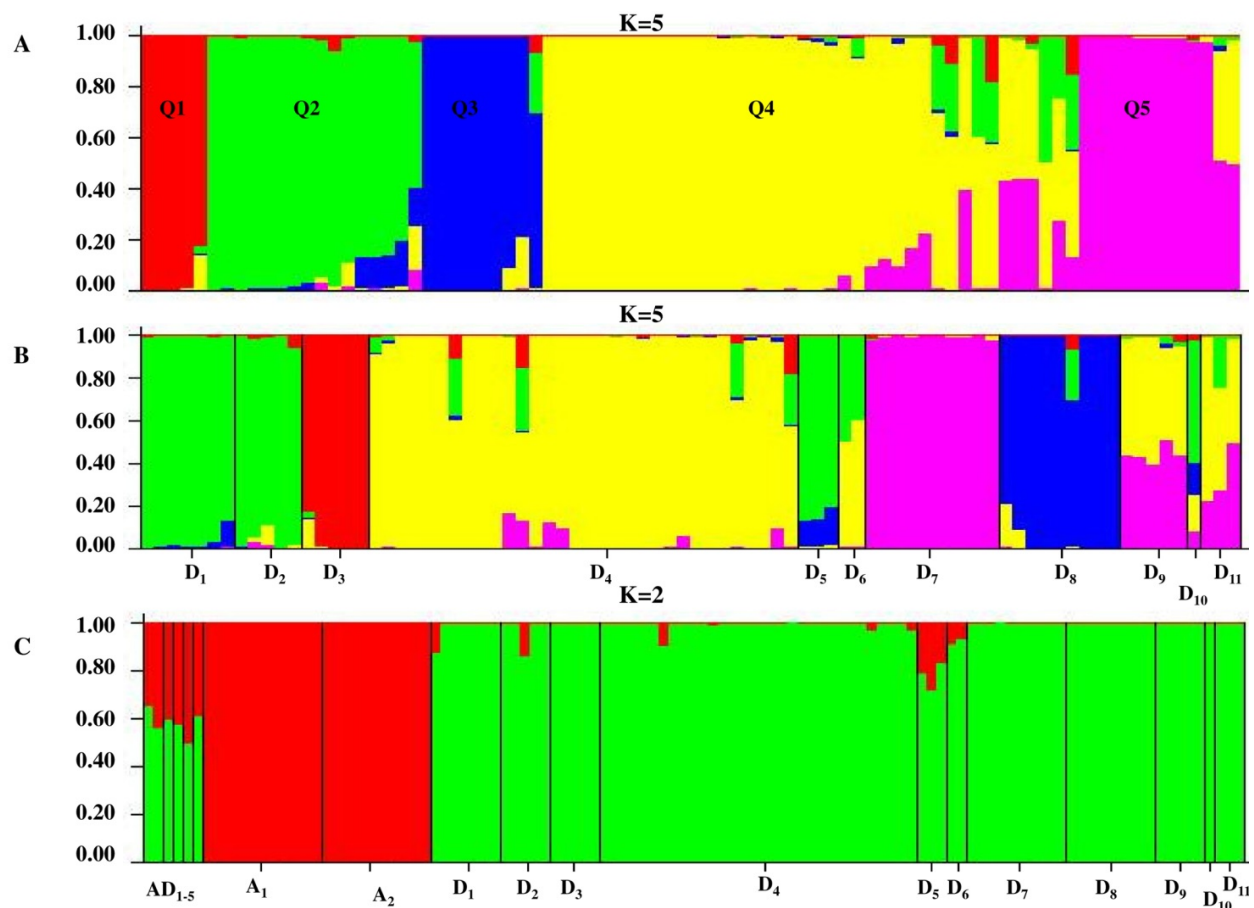

Figure 6. A) Phylogenetic neighbor-joining dendrogram arbitrarily rooted of 111 *Gossypium* accessions representing five allotetraploids ($AD_1 – AD_5$ genomes), 23 Asiatic diploids of the Old World (A_1 and A_2 genomes), and 82 diploids of the New World subgenus *Houzingenia* ($D_1 – D_{11}$ genomes). Phylogenetic tree developed based upon the proportion of alleles between accessions and B) Unrooted neighbor-joining dendrogram of the 111 *Gossypium* accessions. Branch lengths are shown with bootstrap values. Groups of cotton species and accessions specific to each ecotype are color-coded for simplicity. This figure was published in the following article: Ulloa M, Abdurakhmonov IY, Perez-M C, Percy R, Stewart McDJ. Genetic diversity and population structure of cotton (*Gossypium* spp.) of the New World assessed by SSR markers. Botany 2013;91 251-259. License agreement was provided by the NRC Research Press – Copyright Clearance Center.

The phylogenetic analysis grouped all species into distinct phylogenetic groups consistent with genomic origin (Fig. 6). Based on the Wright's F_{ST} index using AMOVA analyses [9] of the data sets for all-genomes and the diploid New World D-genomes accessions, the differentiation among the population groups of the different species was highly significant ($P \leq$ 0.0001). A great deal of total genetic variance was attributed to the difference among and within groups, especially within the *G. aridum* population-groups or ecotypes (Table 2 and Fig. 6) [9]. The analysis clustered the diploids of the New World into six sections with the three bushy types [(*Houzingenia* (D_1 and D_8), *Integrifolia* (D_{3-d}), and *Caducibracteolata* (D_{2-1}, D_{2-2}, and D_{10})] and three arborescent types [*Erioxylum* (US-72, D_4, D_7, D_9, and D_{11}), *Selera* (D_6), and *Austroamericana* (D_5)]. The classification of the formally reported subgenus and species boundaries are well-understood [5,7]. These results are in agreement with other molecular studies [23-25,28]. Also, the statistical analysis of *structure* test was used in this study [through measurements of *ad hoc* (ΔK) quantity of Evanno statistics] to identify real number of K populations for the germplasm accessions (Table 2). The population structure analysis on this study shed light on the emergence and dispersion of the diploids of the New World and agreed with the hypothesis of a rapid radiation of the American diploid cotton linage that took place somewhere in southwestern Mexico, followed by a differentiation-speciation [9,23-25,28]. This radiation might have occurred before the separation of the Baja California peninsula (7-12 million years ago) from the mainland of the country of Mexico [9,28]. The population structure analyses [9] indicated that Baja California peninsula was colonized from two independent lineages, one from the subsection *Intergrifolia* (Q1, D_3-accessions) and the second from the subsection *Caducibracteata* (Q2, D_2-accessions) (Fig. 7). These two species (*G. harknessii*-D_2-accessions and *G. davidsonii*-D_3-accessions) are clearly distinguished by many morphological features: leaves, flowers, seed capsule, pubescence, etc.

The population structure analyses [9] with the geographic distribution and morphology of some of these species also supports the hypothesis that the New World D diploid species may derive from five major lineages (Q1-Q5) that eventually radiated and differentiated about 7-8 million years ago through the country of Mexico. Some species [*G. gossypioides* (D_6 genome), *G. laxum* (D_9 genome), *G. turneri* (D_{10} genome), and *G. schwendimanii* (D_{11} genome)] experienced a more recent differentiation event (Fig 7B). Interspecific gene flow has been recognized as an important evolutionary event in plants. It has also been suggested that improbable interspecific introgression and molecular differentiation may have occurred more often than predicted in angiosperm evolution [9-10,28]. Supra-specific coalescence of some alleles in these species may support the mixed sample-group of the D genome accessions (D_4, D_6, D_9, D_{10}, and D_{11}), experiencing more recent hybridization events.

The phylogenetic analysis grouped all species into distinct phylogenetic groups of the New World cottons (Fig. 8), consistent with genomic origin and classified species of the *Houzingenia* subgenus with the six subsections: subsection *Austroamericana* [*G. raimondii* (D_5)]; subsection *Caducibracteolata* [*G. armourianum* (D_{2-1}), *G. harknessii* (D_{2-2}), and *G. turneri* (D_{10})]; subsection *Houzingenia* [*G. thurberi* (D_1) and *G. trilobum* (D_8)]; subsection *Integrifolia* [*G. davidsonii* (D_{3-d})]; subsection *Erioxylum* [*G. aridum* (D_4), *G. lobatum* (D_7), *G. laxum* (D_9), and *G. schwendimanii* (D_{11})]; and subsection *Selera* [*G. gossypioides* (D_6)] [3,4,7]. In addition, several non-described taxa of

Figure 7. Population structure of cotton (*Gossypium* spp.) accessions: (A) Eighty-two D-genome cotton accessions (D₁ – D₁₁) population structure plots at K=5 grouped by Q-matrix, (B) grouped by genomic origins of the D-genome accessions, and C) 111 *Gossypium* accessions representing five allotetraploids (AD₁ – AD₅ genomes), 23 Asiatic diploids of the Old World (A₁ and A₂ genomes), and 82 diploids of the New World subgenus *Houzingenia* (D₁ – D₁₁ genomes)-All-genome cotton accessions population structure plot at K=2, where Q1 (red) represents diploid A-genome accessions (A₁ and A₂), Q2 (green) represent diploid D-genomes (D₁ – D₁₁), and red and green represents the amalgamation of the tetraploid AD genomes (AD₁ – AD₅). This figure was published in the following article: Ulloa M, Abdurakhmonov IY, Perez-M C, Percy R, Stewart McDJ. Genetic diversity and population structure of cotton (*Gossypium* spp.) of the New World assessed by SSR markers. Botany 2013; 91 251-259. License agreement was provided by the NRC Research Press – Copyright Clearance Center.

the New World diploid *G. aridum* species were found to be distanced from their groups or ecotypes from the states of Nayarit, Guerrero, and Oaxaca (Fig. 8). As mentioned before, *G. gossypioides* has been reported with cryptic repeated genomic recombination during speciation, with conflicting morphological, cytogenetic, and molecular evidence of its phylogenetic affinity to other New World cottons [13]. It has been proposed that *G. gossypioides* might hybridize with an African A-genome and/or extinct taxon based on transfer of repetitive DNA [13]. In the neighbor-joining method, trees are constructed by linking together the two operational taxonomic units or in other words – leaves of the tree, or hypothetical taxonomic units that are the closest mutual "neighbors" [29-30]. The phylogenetic resolution of *G. gossypioides* has been found to be inconsistent because this species has been placed within the New World cotton of the D genome in a basal clade-position of the phylogenetic tree rather
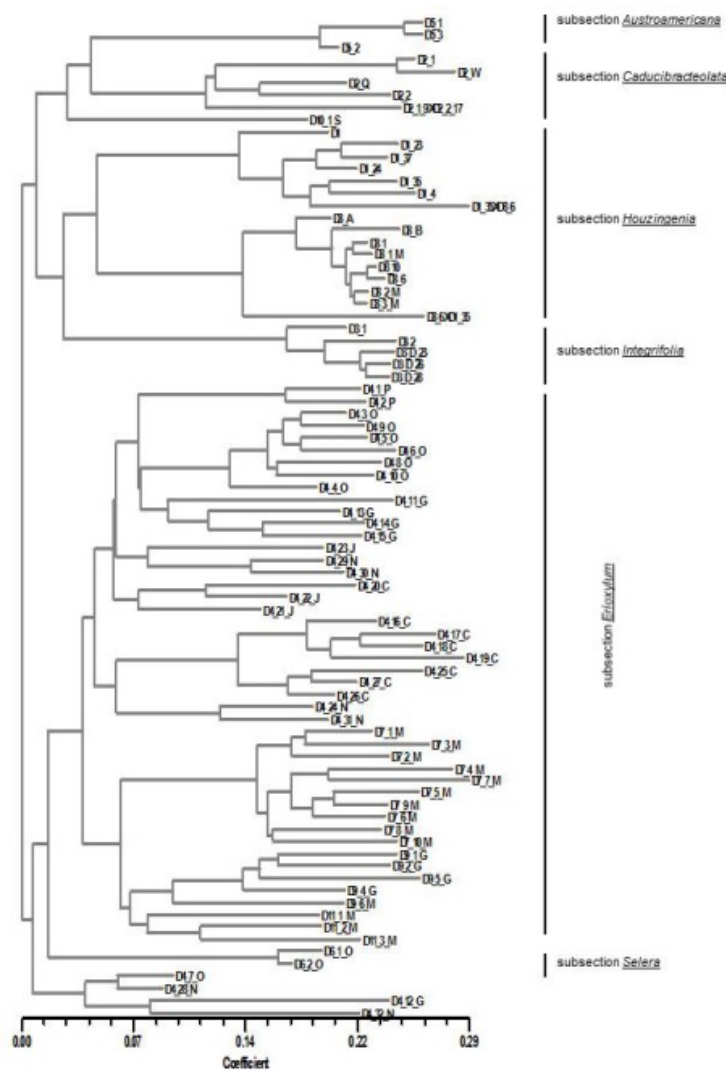
Figure 8. A) Phylogenetic neighbor-joining dendrogram using the midpoint rooted of the 82 diploids of the New World subgenus *Houzingenia* (D$_1$ – D$_{11}$ genomes). Phylogenetic tree developed based upon the proportion of alleles between accessions.

than in the same clade with the previously proposed sister, *G. raimondii* [13]. This proposed relationship between these two species was based on phylogentic studies with chloroplast (cpDNA) genes. In Figure 8, *G. gossypioides* has been placed at the basal clade-position of the arborescent subsection *Erioxylum* while *G. raimondii* subsection *Austroamericana* shared a clade-position with species of the *Caducibracteolata* subsection. The clade-position of these two species may indicate two divergent evolutionary events through introgression or hybridization.

The arborescent subsection *Erioxylum* is among the most distinctive in the genus [31]. However, the sectional-levels of *G. aridum*, as formally reported [5-8], still remain unresolved. SSR markers have proved to be a powerful tool in elucidating genetic relationships and population structure of these accessions [9]. A proposed genetic distance (GD) minimum threshold of 0.20 [9] may be useful to define a new taxon, and a clear relationship among cotton species or

genetically distant geographical accession-ecotypes of *G. aridum.* In addition to US-72 (newly identified taxon) [10,27], five newly collected accessions [D_{4-10-O} (US-41), D_{4-2-P} (US-05), D_{4-12-G} (US-76), D_{4-19-C} (US-122), and D_{4-32-N} (US-150)] from five different ecotypes and states from the country of Mexico were proposed by Ulloa et al [9] to be recognized as new species based on GD. These collected accessions had the larger GD when compared with any other recognized *Gossypium* species of the D genome, GD > 0.28 and GD ≤ 0.41 [9]. Based on the most recent explorations/collections in the country of Mexico [8-9], the existing taxonomic classification of *Gossypium* of the D_4 diploid species made by Fryxell [31] and Fryxell et al [7] needs to be revised.

5. Evolution and review of beneficial genes of the D genomes

The evolution of the *Gossypium* genus started around 10-20 million years ago [32-33]. The initial step of this process might be started with the formation or origin of the American diploids or New World cottons, which may be estimated at around 6.7 million years ago. Following the formation of the diploid cottons was the allopolyploid formation that derived the New World tetraploid cottons around 1-2 million years ago [34], which included *G. hirsutum* and *G. barbadense* cottons. The origin of the allotetraploids is still not well understood. However, it is well established that the allotetraploids combine one genome derived from an A-genome ancestor and a second genome from a D-genome ancestor [9,33-36]. There is no evidence of any A-genome species in the New World, and there is no evidence of any D-genome species outside the New World. There has been considerable speculation over the years as to which D-genome species is the closest living relative of the ancestor of the Dt subgenome of the allotetraploid cottons. Based on molecular data, the best species model of the allotetraploid (AD) Dt subgenome is *G. raimondii*. However, recent discoveries through molecular data also revealed that *G. gossypioides* may be closer than originally thought to *G. raimondii* despite the geographical separation of these two species based on chloroplast (cpDNA) genes [13]. In terms of haploid nuclear DNA content or amounts (1C) the *Gossypium* genomes range from 1 to 3.8 pg=picograms (980 Mbp to 3425 Mbp). The D model genome is smaller with 2C amounts of 1 pg and a haploid length of 980 Mbp while the A-genome diploid nuclear genome contains about 3.8 pg of DNA (2C) and the length of a single copy of the genome is approximately 1860 Mbp. The genome size in the AD tetraploids for the most part is additive with 5.8 pg (2C) and with a haploid length of 3835 Mbp [35,37-40].

As previously established, each genomic designation (A, B, C, D…etc.) represents a functional group of chromosomes that share similar sizes and structures, as well as success of interspecific crosses. These designations also help breeders to find sources of genetic variability for the introgression of beneficial genes into elite cultivars and to determine rates of success of the introgression of these beneficial genes. Within the same designed genome, hybrid chromosomes recombine during meiosis and tend to be fertile. However, crosses made using genomes with different designation with similar basic chromosome number; hybrids are generally infertile with few stable bivalents at meiosis [39]. Breeders first turn to sources of genetic variability in the primary germplasm pool or within the same species, which includes wild or exotic and landrace germplasm of *G. hirsutum* and *G. barbadense*

species. The elite public and private cultivars [*G. hirsutum* (Upland) and *G. barbadense* (Pima)] of these species contain a number of traits that originated in the primary germplasm pool e.g., the blight resistance genes [35], the nectariless trait from *G. tomentosum* [41], root-knot nematode resistance from landraces [42-43], resistance to Fusarium (*Fusarium oxysporum* f.sp. *vasinfectum* Atk. Sny & Hans) and Verticillium (*Verticillium dahlia* Kleb) wilt from landraces of *G. hirsutum* and from *G. darwinii* [44].

The diploid species of the A and D genome belong to the secondary germplasm pool and have contributed to improving Upland and Pima cultivars [4]. Bacterial blight resistance genes from species such as *G. arboreum*, *G. herbaceum* and *G. anomalum* have been introgressed into Upland cultivars [35]. Cytoplasm and restorer factors from *G. harknessii* [45] and *G. trilobum* [46] conditioning cytoplasmic male sterility, and D2 smoothness [47] have also been introgressed into Upland cultivars using these diploid species. Moreover, improvement of fiber quality characteristics or properties has been done via the triple hybrid (*G. hirsutum* x *G. arboreum* x *G. thurberi*) [48]. Introgression of high fiber strength and improvement of fiber quality parameters were obtained using progeny from these hybrid combinations. In addition, similar triple hybrid combinations that include *G. thurberi* [49] and *G. aridum* [50] have provided progeny that have been used to develop resistant germplasm and cultivars to root-knot nematode (*rkn*, *Meloidogyne incognita* Kofoid and White) and reniform nematode (Ren[ari], *Rotylenchulus reniformis* Linford and Oliveira). Resistance to several pests and diseases has been found in diploid cottons. However, in nature, the hybridization of diploid species with allotetraploid (Upland or Pima) species produces sterile hybrids because uneven genome or chromosome basic number and pairing during meiosis. One of the satisfactory mutagenic agents used by the breeders to induce doubling of chromosomes and balance chromosome paring on hybrids is colchicine. The difficulties of obtaining agronomically suitable introgressed progeny are high through this type of interspecific hybridization. The most successful method of introgression has been via hexaploid bridging.

Even though the *Gossypium* species of the D genome are not well known and utilized in cotton improvement and breeding, their significance as great reservoirs of important genes is starting to be noticed and documented. In a comparison quantitative trait loci (QTL) review-study [4], the Dt subgenome exhibits from 32% [4,51] to 57% [52] of QTLs on different chromosomes with QTL effects on different important traits for cotton improvements. These QTLs were located on different chromosomes of the Dt subgenome. And even though the species of the D genome does not produce spinnable fibers, the Dt subgenome of the tetraploid cotton was found to possess QTLs positively affecting fiber quality and morphological traits [53-55] and therefore harboring greater allelic diversity among tetraploid forms. Recently, based on the concept that some diploid species are tolerant to stress and may harbor important genes, a large number of genes were obtained from leaf and root tissues of the diploid *G. aridum* species. Plants of this species were subjected to various salt stresses to examine gene expression and to understand the salt tolerance mechanisms in *Gossypium* [56]. Most of the salt-regulated transcripts were found to be homologous to genes that are known to be associated with salt tolerance e.g., ethylene-responsive transcript factor, aquaporin PIP1, protein kinases (CBL-interacting and mitogen-activated) [56]. New transcriptome data from these plant tissue-species when

eventually compared with available marker-QTL DNA sequence data and/or whole genome sequences will provide new insights into the evolution and expression of genes affecting important traits of cotton. QTL hotspots have been found affecting multiple fiber traits [32,51,57]. DNA sequences of marker-QTLs were found to be contributed by the D genome based on changes in expression of functionally diverse cotton genes [29]. Additional studies using the next generation sequencing (NGS) technology will provide additional information on these unique flowering and fruiting habits (following defoliation in the dry season), and salt and drought resistance mechanisms that allow the D genome species to survive extended periods without rain or stress conditions.

6. Future research of gene discovery and mining of the D genomes

The most widely cultivated cotton species in the world, which is known by various common names (e.g., Acala or Upland cotton, short staple cotton, Mocó cotton, and Cambodia cotton) is *G. hirsutum* [4-5,8,12]. Recent advances in genomics have provided considerable information regarding the discovery and expression of genes controlling important crop traits. In the future, the new generation of cotton breeders will have the opportunity to benefit from the vast information generated by NGS on genomic research. This information could be used to improve existing tools such as MAS for molecular breeding or to develop new tools to locate and characterize germplasm and cultivars or gene-sequences of DNA encoding proteins that controlled their expression of important traits, developmentally and temporally.

With the decrease in sequencing cost by NGS technology, it has been possible to obtain large numbers of base pairs of DNA sequences for identifying polymorphisms and directly mapping genes responsible for important traits through direct whole genome sequencing [58-59]. Molecular markers are being continuously developed, which will allow cotton geneticists to sample all regions of the cotton genome [4,9,32,60-61]. Plant breeders find molecular markers useful as a selection tool in monitoring alien genome introgression in cotton breeding programs [62-63]. Alien genome has the potential to increase genetic variability for economically valuable traits in cotton cultivars. The process of introgression of alien genes/genomes is not easy but clearly increases the amount of genetic diversity available for selection because it is likely that many useful alleles are to be found outside the current cultivated gene pools. Even when the inferred gene is yet to be located, sequenced, and characterized, molecular breeders could use these natural DNA sequence (SNP – single nucleotide polymorphism) to identify differences among germplasm and breeding lines, and applying traditional genetic analyses to infer genes for MAS. In addition, the efficacy of transgenic technology is entirely dependent on gene discovery. The functional genes identified with NGS would be important resources to improve the cotton crop through transgenic technology.

The understanding of the cotton genome is complex, especially the evolution and function of the major cultivated species, e.g., *G. hirsutum*. This complexity arises from the joint presence of the two subgenomes (At and Dt) in its nucleus [60]. The complete sequence of the allotetraploid genomes, including *G. hirsutum*, is still not completed. However, recently the genome

sequence of the best model of the Dt subgenome *(G. raimondii)* was published [32,61]. *G. raimondii* was found to be 47% (around 350 Mb) of euchromatin, spanning 2,059 centiMorgan (cM) and 53% (around 390 Mb) of heterochromatin, spanning a repeat-rich of 186 cM. Transposable-elements accounted for 61% in which 53% were long-terminal-repeats (LTRs) retrotransposons. It was reported that the *G. raimondii* genome contains around 37, 505 assembled genes with 77,267 protein-coding annotated transcripts [31].

Increasing our knowledge and understanding of how cotton (*Gossypium* spp.) can sustain yield under drought and attack of pathogens is essential for sustained profitability and long-term survival of the cotton industry. Researchers and breeders are working to develop sources of germplasm that can improve water use efficiency (WUE), drought and extreme heat tolerance. An emergent concept from genomic studies is that different regulatory networks related to plant stress may be interconnected. For example, in most situations heat stress and drought stress are thought to be linked, and often resistance to pathogens is comprised by abiotic stress factors [64]. Improving cotton productivity in stress environments calls for the understanding of many traits involved in different system-level interactions. The integration of these plant responses and interactions with quantitative phenotypes is complex and requires the use of new approaches and technologies. These new NGS data and analyses will provide new information about the utility of the newly published *G. raimondii* genome sequence to target traits of interest in the allopolyploid species. When new NGS data of the future sequenced genomes and the genome saturation get combined with quantitative genetic analyses, cotton breeders finally will have the tools they need to identify the location of the genes (quantitative trait loci or QTLs) conditioning the expression of critical agronomic traits, such as yield, drought tolerance and water use efficiency [65-66]. If *in situ* diversity of the New World cottons is severely eroded, then current and additional accessions in International Collections and in the USDA Cotton Germplasm Collection assume a highly significant role in preservation of the diversity previously residing in these D genome species. The key to increasing genetic diversity among cultivated cottons is to continue collecting, evaluating and utilizing many different cotton germplasm, including diploid species of the *Gossypium* genus. The diploid D genome cottons (*Gossypium* spp.) of the New World are part of a great reservoir of important genes for improving fiber quality, pest and disease resistance, and drought and salt tolerance in the modern cultivated Upland/Acala (*G. hirsutum*) and Pima or Sea Island (*G. barbadense*) cottons.

Acknowledgements

and does not imply recommendation or endorsement by the U. S. Department of Agriculture. The U. S. Department of Agriculture is an equal opportunity provider and employer.

Author details

Mauricio Ulloa[*]

Address all correspondence to: Mauricio.Ulloa@ars.usda.gov

USDA-ARS, SPA, Cropping Systems Research Laboratory, Plant Stress and Germplasm Development, Lubbock, TX, USA

References

[1] Abdurakhmonov IY, Buriev ZT, Shermatov SE, Abdullaev AA, Urmonov K, Kushanov F, Egamberdiev SS, Shapulatov U, Abdukarimov A, Saha S, Jenkins JN, Kohel RJ, Yu JZ, Pepper AE, Kumpatala S, Ulloa M. Genetic diversity in *Gossypium* genus. In: Çalişkan M.(ed.) Genetic Diversity in Plants. New York, NY: InTech; 2012; p313-338.

[2] Fryxell PA. A revised taxonomic interpretation of *Gossypium* L. (Malvaceae). Rheedea 1992;2 108-165.

[3] Percival AE, Wendel JF, Stewart JM. Taxonomy and Germplasm Resources. In: Smith CW and Cothren JT (eds.) Cotton: Origin, history, technology, and production. New York: John Wiley & Sons; 1999; p. 33-64.

[4] Ulloa M, Brubaker C, Chee P. Cotton. In: Kole C. (ed.) Genome Mapping & Molecular Breeding. Vol. 6: Technical Crops. Heidelberg, Berlin, New York, Tokyo: Springer; 2007; p1-49.

[5] Fryxell PA. The Natural History of the Cotton Tribe. College Station and London: Texas A&M University Press; 1979.

[6] Kantartzi SK, Ulloa M, Sacks E, Stewart JM. Assessing Genetic Diversity in *Gossypium arboreum* L. Cultivars using Genomic and EST-derived Microsatellites. Genetica 2009; 136:141-147.

[7] Fryxell PA, Craven LA, Stewart JM. A revision of *Gossypium* sect.*Grandicalyx* (Malvaceae), including the description of six new species. Systematic Botany. 1992; 17:91-114.

[8] Ulloa, M, Stewart JM, Garcia-C EA, Godoy-A A, Gaytán-M A, Acosta-N S. Cotton genetic resources in the western states of Mexico: in situ conservation status and germplasm collection for ex situ preservation. Genet. Resour. Crop Evol; 2006. 53: 653-668.

[9] Ulloa M, Abdurakhmonov IY, Perez-M C, Percy R, Stewart McDJ. Genetic diversity and population structure of cotton (*Gossypium* spp.) of the New World assessed by SSR markers. Botany 2013; 91:251-259.

[10] Alvarez I, Wendel JF. Cryptic interspecific introgression and genetic differentiation within *Gossypium aridum* (Malvaceae) and its relatives. Evolution 2006; 60:505-517.

[11] Feng C, Ulloa M, Perez-M C, Stewart, JM. Distribution and molecular diversity of arborescent *Gossypium* species. Botany; 2011; 89:615-624

[12] Fryxell PA, Taxonomy and Germplasm Resources, In: Kohel RJ and Lewis CF (eds) Cotton. Agronomy No. 24, ASA, CSSA & SSSA. 1984; p. 27-58.

[13] Cronn RC, Small RL, Haselkorn T, Wendel JF. Duplicated genes evolve independently after polyploid formation in cotton. 2003; Evolution: 2475-2489.

[14] Ram SG, Thiruvengadam V, Vinod KK. Genetic diversity among cultivars, landraces and wild relatives of rice as revealed by microsatellite markers. J. Appl. Genet. 2007; 48:337-345.

[15] Vigouroux Y, Glaubitz JC, Matsuoka Y, Goodman MM, Sanchez GJ, Doebley J. Population structure and genetic diversity of New World maize races assessed by DNA microsatellites. Am. J. Bot. 2008; 95:1240-1253.

[16] Wendel JF, Schnabel A, Seelanan T. An unusual ribosomal DNA sequence from *Gossypium gossypioides* reveals ancient, cryptic intergenomic introgression. Mol. Phylogen. Evol. 1995; 4:298-313.

[17] Buckler ES, Ippolito A, Holtsford TP. The Evolution of ribosomal DNA: divergent paralogues and phylogenetic implications. Genetics. 1997; 145:821-832.

[18] Pillay M, Myers GO. Genetic diversity in cotton assessed by variation in ribosomal RNA genes and AFLP markers. Crop Sci. 1999; 39:1881-1886.

[19] Small RL, Ryburn JA, Cronn RC, Seelanan T, Wendel JF. The tortoise and the hare: Choosing between noncoding plastome and nuclear Adh sequences for phylogeny reconstruction of a recently diverged plant group. Am. J. Bot. 1998; 85:1301-1315.

[20] Zhao X, Si Y, Hanson RE, Crane CF, Price HJ, Stelly DM, Wendel JF, Paterson AH. Dispersed repetitive DNA has spread to new genomes since polyploidy formation in cotton. Genome Research. 1998; 8:479-492.

[21] Hanson RE, Zhao XP, Islam-Faridi MN, Paterson AH, Zwick MS, Crane CF, McKnight TD, Stelly DM, Price HJ. Evolution of interspersed repetitive elements in *Gossypium* (Malvaceae). Am. J. Bot. 1998; 85:1364-1368.

[22] Hanson RE, Islam-Faridi MH, Crane CF, Zwick MS, Czeschin DG, Wendel JF, McKnight TD, Price HJ, Stelly DM. Ty1-copia-retrotransposon behavior in a poly-ploid cotton. Chromosome Research 1999; 8:73-76.

[23] Cronn RC, Small RL, Wendel JF. Duplicated genes evolve independently after poly-ploid formation in cotton. Proc. Natl. Acad. Sci. 1999; 96:14406-14411.

[24] Small RL, Wendel JF. Phylogeny, duplication and intraspecific variation of Adh se-quences in new world cotton (*Gossypium* L., Malvaceae). Mol. Phylogen. Evol. 2000a; 16:73-84.

[25] Small RL, Wendel JF. Copy number liability and evolutionary dynamics of the Adh gene family in diploid and tetraploid cotton (*Gossypium*). Genetics. 2000b; 155:1913-1926

[26] Liu Q, Brubaker CL, Green AG, Marshall DR, Sharp PJ. Singh S.P. Evolution of the FAD2-1 fatty acid desaturase 5' UTR intron and the molecular systematics of *Gossypi-um* (Malvaceae). Am. J. Bot. 2001; 88:92-102.

[27] Wendel JF, Cronn RC, Johnston JS, Price HJ. Feast and famine in plant genomes. Ge-netica. 2002; 115:37-47.

[28] Álvarez I, Cronn R, Wendel JF. Phylogeny of the New World diploid cottons (*Gos-sypium* L., Malvaceae) based on sequences of three low-copy nuclear genes. Plant Systematics and Evolution. 2005; 252:199-214.

[29] Rohlf FJ. NTSYS pc: Numerical taxonomy system, Ver: 2.01. 2002; Exter publishing, Ltd. Setauket, NY.

[30] Saitou M, Nei N. The neighbor joining method: a new method for reconstructing phylogenetic trees. 1987; Mol Biol Evol 4: 406-425.

[31] Fryxell PA. A revised taxonomic interpretation of *Gossypium* L. (Malvaceae). Rhee-dea. 1992; 2: 108-165.

[32] Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, Jin D, Llewellyn D, Show-maker KC, Shu S, Udall J, Yoo M, Byers R, Chen W, Faigenboim AD, Duke MV, Gong L, Grimwood J, Grover C, Grupp K, Hu G, Lee T, Li J, Lin L, Liu T, Marler BS, Page JT, Roberts AW, Romanel E, Sanders WS, Szadkowski E, Tan X, Tang H, Xu C, Wang J, Wang Z, Zhang D, Zhang L, Ashrafi H, Bedon F, Bowers JE, Brubaker CL, Chee PW, Das S, Gingle AR, Haigler CH, Harker D, Hoffmann LV, Hovav R, Jones DC, Lemke C, Mansoor S, Rahman M, Rainville LN, Rambani A, Reddy UK, Rong J, Saranga Y, Scheffler BE, Scheffler JA, Stelly DM, Triplett BA, Van Deynze A, Vaslin MFS, Waghmare VN, Walford SA, Wright RJ, Zaki EA, Zhang T, Dennis ES, Mayer KFX, Peterson DG, Rokhsar DS, Wang X, Schmutz J. Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. Nature 2012; 492: 423–427.

[33] Wendel JF, Cronn RC. Polyploidy and the evolutionary history of cotton. Adv. Agro, Academic Press. 2003; 78:139-186.

[34] Senchina DS, Alvarez I, Cronn R, Liu B, Rong J, Noyes RD, Paterson AH, Wing RA, Wilkins TA, Wendel JF. Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. Mol. Biol. Evol. 2003; 20:633-643.

[35] Endrizzi JE, Turcotte EL, Kohel RJ. Genetics, cytogenetics and evolution of *Gossypium.* Adv. Genet. 1985; 23:271-375.

[36] Brubaker CL, Bourland FM, Wendel JF. The origin and domestication of cotton. In Cotton: Origin, History, Technology and Production. Smith CW and Cothren JT (eds). Wiley, New York. 1999; p. 3-31.

[37] Edward GA, Endrizzi JE, Stein R. Genomic DNA content and chromosome organization in *Gossypium*. Chromosoma. 1974; 47:309-326.

[38] Bennett MD, Smith JB, Heslop-Harrison JS. Nuclear DNA amounts in angiosperms. Proc. R. Soc. Lond. B. 1982; 216:179-199.

[39] Stewart JM, Potential for crop improvement with exotic germplasm and genetic engineering. In Challenging the future: Proceedings of the World Cotton Research Conference. Constable GA and Forrester NW (eds). CSIRO Melbourne. 1995; p.313-327.

[40] Hendrix B, Stewart JM. Estimation of the nuclear DNA content of *Gossypium* species. Ann. Bot. 2005; 95:789-797.

[41] Meyer JR, Meyer VG. Origin and inheritance of nectariless cotton. Crop Sci. 1961; 1:167-169.

[42] Shepherd RL. Breeding root-knot-resistant *Gossypium hirsutum* L. using a resistant wild *G. barbadense* L. Crop Sci 1974; 14:687-691.

[43] Shepherd RL. Registration of Auburn 634. Crop Sci. 1982; 22:642.

[44] Bell AA. Cotton protection practices in the USA and world. Section B: Diseases. In: Kohel RJ, Lewis CF (eds) Cotton, Agronomy Monograph 24. ASSA, Madison, WI, USA. 1984; pp 288-309.

[45] Meyers VG. Male sterility from *Gossypium harknessii*. J Hered. 1975; 66:23-27.

[46] Stewart MJ. A new cytoplasmic male sterile and restorer. Proceedings Beltwide Cotton Conferences; USA. 1992; p610

[47] Meyer JR. Origin and inheritance of D2 smoothness in upland cotton. Crop Sci. 1957; 48:249-250.

[48] Harrell SC, Culp TW. Registration of Pee Dee 0259 and Pee Dee 2165 germplasm lines of cotton (Reg No GP39 and GP40). Crop Sci.1979; 19:418.

[49] Ulloa M, Wang C, Roberts PA. Gene action analysis by inheritance and quantitative trait loci mapping of resistance to root-knot nematodes in cotton. Plant Breeding. 2010; 129:541-550.

[50] Romano GB, Sacks EJ, Stetina SR, Robinson AF, Fang DD, Gutierrez OA, Scheffler JA. Identification and genomic location of a reniform nematode (*Rotylenchulus reniformis*) resistance locus (*Ren*[aris]) introgressed from *Gossypium aridum* into upland cotton (*G. hirsutum*). Theor Appl Genet 2009; 120:139-150.

[51] Ulloa M. Saha S, Jenkins JN, Meredith WR Jr, McCarty JC, Stelly MD. Chromosomal assignment of RFLP linkage groups harboring important QTLs on an intraspecific cotton (*Gossypium hirsutum* L.) joinmap. J. Hered. 2005; 96132-144.

[52] Paterson AH, Saranga Y, Menz M, Jiang C, Wright RJ. QTL analysis of genotype x environmental interactions affecting cotton fiber quality. Theor Appl Genet. 2003; 106:384-396.

[53] Reinisch AJ, Dong JM, Brubaker CL, Wendel JF, Paterson AH. A detailed RFLP map of cotton, *Gossypium hirsutum x Gossypium barbadense*: chromosome organization and evolution in a disomic polyploid genome. Genetics. 1994; 138:829-847.

[54] Rong J-K, 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. A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). Genetics. 2004; 166:389-417.

[55] Wright RJ, Thaxton P, El-Zik K, Peterson AH. D-subgenome bias of *Xcm* resistance genes in tetraploid *Gossypium* (cotton) suggest that polyploidy formation has created novel avenues for evolution. Genetics. 1998; 149:1978-1996.

[56] Xu P, Liu Z, Fan X, Gao J, Zhang X, Zhang X, Shen X. De novo transcriptome sequencing and comparative analysis of differentially expressed genes in *Gossypium aridum* under salt stress. Gene. 2013; 525:26-34.

[57] 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. Meta-analysis of polyploid cotton QTL shows unequal contributions of subgenomes to a complex network of genes and gene clusters implicated in lint fiber development. 2007; Genetics 176: 2577-2588.

[58] Abe A, Kosugi S, Yoshida K, Natsume S, Takagi H, Kanzaki H, Matsumura H, Mitsuoka C, Tamiru M, Innan H, Cano L, Kamoun S, Terauchi R.. Genome sequencing reveals agronomically important loci in rice using MutMap. Nat. Biotechnol. 2012; 30:174-178.

[59] Austin RS, Vidaurre D, Stamatiou G, Breit R, Provart NJ, Bonetta D, Zhang J, Fung P, Gong Y, Wang PW, McCourt P, Guttman DS. Next-generation mapping of Arabidopsis genes. Plant J. 2011; 67:715-725.

[60] Page JT, Huynh MD, Liechty ZS, Grupp K, Stelly D, Hulse AM, Ashrafi H, Van-Deynze A, Wendel JF, Udall JA. Insights into the evolution of the cotton diploids and polyploids from whole-genome re-sequencing. G3, 2013; 3: 1809-1818.

[61] Wang K, Wang Z, Li F, Ye W, Wang J, Song G, Yue Z, Cong L, Shang H, Zhu S, Zou C, Li Q, Yuan Y, Lu C, Wei H, Gou C, Zheng Z, Yin Y, Zhang X, Liu K, Wang B, Song C, Shi N, Kohel RJ, Percy RG, Yu JZ, Zhu Y-X, Wang J, Yu S. The draft genome of a diploid cotton *Gossypium raimondii*. Nature Genetics. 2012; 44(10) 1098-1104.

[62] Yu JZ, Kohel RJ, Fang DD, Cho J, Van Deynze A, Ulloa M, Hoffman SM, Pepper AE, Stelly DM, Jenkins JN, Saha S, Kumpatla SP, Shah MR, Hugie WV, Percy RG. A high-density SSR and SNP genetic map of the tetraploid cotton genome. G3: Genes, Genomes, Genetics 2012; 2:43-58.

[63] Ulloa M, Percy R, Hutmacher RB, Zhang J. The future of cotton breeding in the Western United States. J. Cot. Sci. 2009; 4 246-255.

[64] Zhang L, Li Y, Lu W, Meng F, Wu CA, Guo X. Cotton GhMKK5 affects disease resistance, induces HR-like cell death, and reduces the tolerance to salt and drought stress in transgenic *Nicotiana benthamiana*. J. Exp Bot. 2012; 63:3935-3951.

[65] Saranga Y, Menz M, Jiang C, Wright RJ, Yakir D, Paterson AH. Genomic Dissection of genotype x environment interactions conferring adaptation of cotton to arid conditions. Genome Res. 2001; 11:1988-1995.

[66] Burke JJ, Ulloa M. Initial characterization of a Stoneville 474 x Phytogen 72 recombinant inbred population. In: Boyd S, Huffman M, Robertson B. (eds.): Proceedings of the Beltwide Cotton Conferences, 7-10 January 2013, San Antonio, TX, USA. Memphis: National Cotton Council.