Genetic diversity and population structure of cotton (Gossypium spp.) of the New World assessed by SSR markers

Article *in* Botany · January 2013

DOI: 10.1139/cjb-2012-0192

CITATIONS
31

READS
296

5 authors, including:

Mauricio Ulloa
University of Concepción
119 PUBLICATIONS **2,908** CITATIONS
SEE PROFILE

Ibrokhim Y Abdurakhmonov
Ministry of Innovational Development
130 PUBLICATIONS **2,044** CITATIONS
SEE PROFILE

Claudia Pérez Mendoza
Colegio de Postgraduados
25 PUBLICATIONS **81** CITATIONS
SEE PROFILE

Richard Greeer Percy
Agricultural Research Service
141 PUBLICATIONS **5,681** CITATIONS
SEE PROFILE

NRC Research Press

ARTICLE

Genetic diversity and population structure of cotton (*Gossypium* spp.) of the New World assessed by SSR markers

Mauricio Ulloa, Ibrokhim Y. Abdurakhmonov, Claudia Perez-M., Richard Percy, and James McD. Stewart

Abstract: A global analysis of cotton (*Gossypium* spp.) genetic diversity is the first step to understanding its geographical distribution, dissemination, genetic relatedness, and population structure. To assess the genetic diversity and population structure in *Gossypium* species, 111 cotton 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) species were assessed using simple sequence repeats (SSR) markers with wide genome coverage. The mean genetic distance (GD) between the two most important New World tetraploid cottons (Upland (*Gossypium hirsutum* L.) and Pima (*Gossypium barbadense* L.)) was 0.39. Among the three shrub type sections (*Houzingenia*, *Integrifolia*, and *Caducibracteolata*) and three arborescent sections (*Erioxylum*, *Selera*, and *Austroamericana*), the GD ranged between 0.19 and 0.41. Phylogenetic analyses clustered all species into distinct phylogenetic groups, which were consistent with genomic origin, evolutionary history, and geographic distribution or ecotypes of these accessions, suggesting the existence of clear structured strata. With all of the genomes, the highest statistical analysis of Structure test through measurements of ad hoc (ΔK) occurred at $K = 2$, with group Q1 with the Old World diploid A genomes and with group Q2 with all the New World diploids of the D genome. AD genome accessions shared nearly equal alleles from both Q1 and Q2 groups. With all of the diploids of the New World D genomes, the highest value of ΔK occurred at $K = 5$. These results are consistent with the fundamental knowledge of tetraploid AD-genome formation and the rapid radiation of the American diploid cotton linage that took place somewhere in southwestern Mexico, followed by a differentiation–speciation during angiosperm evolution. In addition, SSR markers provide an alternative solution for distinguishing phylogenetic relationships between accessions of different ecotypes and for elucidating population structure of cottons of the New World.

Key words: Gossypium, cotton, D genome, molecular diversity, genetic resource, phylogenetic relationship, germplasm improvement.

Résumé : Une analyse globale de la diversité génétique du coton (*Gossypium* spp.) constitue un premier pas pour en comprendre la distribution, la dissémination, la parenté génétique et la structure des populations. Pour évaluer la diversité génétique et la structure des populations, en utilisant les marqueurs à large couverture génomique des espèces, les auteurs ont étudié 111 accessions représentant cinq espèces allotétraploïdes (génomes AD_1–AD_5), 23 diploïdes asiatique de l'Ancien Monde (génomes A_1 and A_2), et 82 diploïdes du Nouveau Monde du sous-genre Houzingenia (génomes D_1–D_{11}). La distance génétique moyenne (DG) entre les deux cotons tétraploïdes les plus importants du Nouveau Monde (hautes terres (*Gossypium hirsutum* L.) et Pima (*Gossypium barbadense* L.) est de 0.39. Parmi les trois sections arbustives (Houzingenia, Integrifolia, et Caducibracteolata) et les trois sections arborescentes (Erioxylum, Selera, et Austroamericana), la DG va de 0.19 à 0.41. Les analyses phylogénétiques regroupent toutes ces espèces en groupes phylogénétiques distincts, ce qui est congru avec l'origine génomique, l'histoire évolutive et la distribution géographique ou les écotypes de ces accessions, ce qui suggère l'existence d'une nette structure stratifiée. Pour l'ensemble des génomes, l'analyse statistique de structure la plus élevée de la mesure ad hoc (ΔK) survient à $K = 2$), dans le groupe Q1 chez les génomes diploïdes A de l'Ancien Monde et dans le groupe Q2 chez les diploïdes du génome D de l'Ancien Monde. Les accessions du génome AD partagent presque également les allèles à la fois des groupes Q1 et Q2. Avec l'ensemble des génomes diploïdes D du Nouveau Monde, la plus haute valeur de ΔK survient à $K = 5$. Ces résultats sont congrus avec la connaissance fondamentale de la formation des génomes tétraploïdes AD et la rapide radiation de la lignée diploïde américaine ayant eu lieu quelque part dans le sud-ouest du Mexique, suivie d'une différenciation–spéciation au cours de l'évolution des angiospermes. De plus, les marqueurs SSR fournissent une solution alternative pour distingue les relations phylogénétiques entre les accessions de différents écotypes et pour élucider la structure des populations de coton du Nouveau Monde. [Traduit par la Rédaction]

Mots-clés : *Gossypium*, coton, génome D, diversité moléculaire, ressource génétique, relations phylogénétiques, amélioration des germplasmes.

Received 10 August 2012. Accepted 5 December 2012.

M. Ulloa. U.S. Department of Agriculture – Agricultural Research Service, Southern Plains Area, Cropping Systems Research Laboratory, Plant Stress and Germplasm Development Research, 3810 4th Street, Lubbock, TX 79415, USA.
I.Y. Abdurakhmonov. The Center of Genomics and Bioinformatics, Academy of Sciences of Uzbekistan, Ministry of Agriculture and Water Resources,"Uzpakhtasanoat" Association, Tashkent, Republic of Uzbekistan.
C. Perez-M. Campo Experimental Iguala, Centro de Investigaciones Pacific sur-INIFAP, Iguala, Gro., Mexico.
R. Percy. U.S. Department of Agriculture – Agricultural Research Service, Southern Plains Area, Crop Germplasm Research. Unit, College Station, TX 79415, USA.
J.McD. Stewart. University of Arkansas, Department of Crop, Soil, and Environmental Sciences, Fayetteville, AR 72701, USA.
Corresponding author: Mauricio Ulloa (e-mail: mauricio.ulloa@ars.usda.gov).

Published at www.nrcresearchpress.com/cjb on 21 January 2013.

Introduction

The allotetraploid cotton (*Gossypium* spp.) produces the world's leading natural fiber and is the second most important oilseed crop. New World tetraploid cottons (Upland, *Gossypium hirsutum* L. (AD_1); Pima, *Gossypium barbadense* L. (AD_2)) dominate world production (98%), whereas Old World Asiatic cottons, *Gossypium herbaceum* L. (A_1) and *Gossypium arboreum* L. (A_2) still remain primarily for nonindustrial consumption in India and Asian adjacent countries (Ulloa et al. 2007; Kantartzi et al. 2009). Even though Upland and Pima cottons dominate world production, the genetic diversity of Upland in the United States (Abdalla et al. 2001; Wallace et al. 2009), China (Xu et al. 2002), Australia (Multani and Lyon 1995), Pakistan (Rahman et al. 2002), and Greece (Linos et al. 2002) is very low. This low genetic diversity is the result of human domestication and breeding of all major crops used in modern agriculture. During domestication, cultivated plants underwent bottleneck selection (Tanksley and McCouch 1997) that generally reduced their genetic diversity relative to their wild ancestors. This low or narrow genetic diversity in cotton cultivars has been suggested as a contributor to an apparent plateau in breeding progress (Meredith 1992; Ulloa 2006), and potentially could limit future cotton improvement (Ulloa et al. 2009).

One solution to the narrow genetic diversity of cultivated plants is to collect, evaluate, and utilize a broader range of cotton germplasm, with special emphasis on diploid species of the *Gossypium* genus (Ulloa et al. 2007; Abdurakhmonov et al. 2011). Species of *Gossypium* are donors of important genes for cotton improvement (Ulloa et al. 2006). 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 (also called linkage drag) (Tanksley and McCouch 1997; Ulloa et al. 2007). This linkage drag has limited the use of such germplasm. However, continuing the introduction of genetic diversity into cultivated plants is important for reducing crop vulnerability (Ulloa et al. 2009). A global analysis of New World cotton genetic diversity is the first step to understanding its geographical distribution, dissemination, genetic relatedness, and the population structure of these species of the *Gossypium* genus.

The amplitude of genetic diversity of cotton is exclusively wide with disperse geographic and ecological niches (Abdurakhmonov 2007; Abdurakhmonov et al. 2011). The *Gossypium* genus of the Malvaceae family contains more than 45 diploid species and five well documented 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 the following designations: AD, A, B, C, D, E, F, G, and K (Percival et al. 1999). The New World AD tetraploid cotton species are large shrubs to trees (Fryxell 1979, 1992; Percival et al. 1999). The center of morphological diversity for *G. hirsutum* (AD_1) is Mesoamerica (southern Mexico–Guatemala), but the indigenous range includes the Caribbean, northern South America, and some Pacific Islands. *Gossypium barbadense* (AD_2) is indigenous to South America but does extend into Mesoamerica and the Caribbean. *Gossypium tomentosum* Nuttall *ex* Seemann (AD_3) is indigenous to the Hawaiian Islands. *Gossypium mustelinum* Miers *ex* Watt (AD_4) is indigenous to northeastern Brazil, and *Gossypium darwinii* Watt (AD_5) is indigenous to the Galápagos Islands (Percival et al. 1999). Only two of the Old World diploid species (*G. arboreum* and *G. herbaceum*) and two of the New World tetraploid cottons (Upland and Pima) are cultivated (Kantartzi et al. 2009).

The New World diploid *Gossypium* comprise 14 species (one nondescribed taxon US-72, Ulloa et al. 2006; Álvarez and Wendel 2006; Feng et al. 2011) from the D genome (Fryxell 1992; Ulloa et al. 2006). 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 extend northward into Arizona. Five species are adapted to the desert environments of Baja California and Sonora (*Gossypium armourianum* Kearney ($D_{2\text{-}1}$), *Gossypium harknessii* Brandegee ($D_{2\text{-}2}$), and *Gossypium davidsonii* Kellogg ($D_{3\text{-}d}$)) and northwestern mainland Mexico (*Gossypium turneri* Fryxell (D_{10}) and *Gossypium thurberi* Todaro (D_1)). An additional seven species (*G.* sp. US-72, *Gossypium aridum* (Rose & Standley) Skovsted (D_4), *Gossypium lobatum* Gentry (D_7), *Gossypium laxum* Phillips (D_9), *Gossypium schwendimanii* Fryx. & Koch (D_{11}), *Gossypium gossypioides* (Ulbrich) Standley (D_6), and *Gossypium trilobum* (Mociño & Sessé ex DC.) Skovsted (D_8)) 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). Two other species have disjointed distributions, as follows: *Gossypium raimondii* Ulbrich (D_5) is endemic to Peru, whereas *Gossypium klotzschianum* Andersson ($D_{3\text{-}k}$) is found in the Galápagos Islands. The D-genome species (subgenus *Houzingenia*) are classified into the following six sections: section *Houzingenia* Fryxell (D_1 and D_8); section *Integrifolia* Todaro ($D_{3\text{-}d}$ and $D_{3\text{-}k}$); section *Caducibracteolata* Mauer ($D_{2\text{-}1}$, $D_{2\text{-}2}$, and D_{10}); section *Erioxylum* Rose & Standley (US-72, D_4, D_7, D_9, and D_{11}); section *Selera* (Ulbrich) Fryxell (D_6); and section *Austroamericana* Fryxell (D_5) (Percival et al. 1999). Although none of these D-genome diploid species produce long or commercially useful fibers, the D genome is one of the parental lineages of the allotetraploid modern cultivated cotton, Upland and Pima (Ulloa et al. 2007).

The most recent taxonomic classification *Gossypium* species was made by Fryxell (1992) and Fryxell et al. (1992). However, traditional taxonomy based on morphology (plant canopy, plant height, leaf and capsule shapes, flowers and petal spots, seed size, etc.) may not distinguish two species with intermediate phenotypes. Molecular methods provide an alternative solution of resolving poorly defined morphological differences between two species. In other crops such as rice (Ram et al. 2007) and maize (Vigouroux et al. 2008), molecular markers such as microsatellites or simple sequence repeats (SSR) have been used to reveal genetic diversity and to distinguish poorly defined differences between species or wild relatives. In cotton, numerous molecular analyses have been performed. *Gossypium* species have been evaluated by internal transcribed spacer (ITS) of ribosomal DNA (Wendel et al. 1995; Pillay and Myers 1999), ribosomal DNA (Wendel et al. 1995; Buckler et al. 1997), and chloroplast DNA (Small et al. 1998), or by repetitive DNA (Zhao et al. 1998; Hanson et al. 1998, 1999), a few loci, such as the *Adh* gene (Small et al. 1998; Cronn et al. 1999; Small and Wendel 2000*a*, 2000*b*), *FAD2-1* gene (Liu et al. 2001), and *Ces* A1 gene (Wendel et al. 2002). More recently, the phylogeny of the New World diploid *Gossypium* was analyzed based on the sequences of three low-copy nuclear genes (Álvarez et al. 2005). Phylogenetic relationships among species of the D genome still remain unclear despite previous molecular studies (Feng et al. 2011). Relationships are unclear in sympatric or closely related species, and marker analyses are needed to resolve local population structures.

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. Recently, the United States Department of Agriculture and the Mexican Instituto Nacional de Investigaciónes Forestales Agricolas y Pecuarias (INIFAP) jointly sponsored *Gossypium* germplasm collection trips by United States 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).

In 2011, Feng et al. (2011) examined molecular diversity and phylogenetic relationships among 33 accessions of arborescent *Gossypium* including 23 of *G. aridum*. These accessions are the most

Table 1. Summary of the 111 accessions used to investigate genetic diversity and population structure of the New World cottons (*Gossypium* spp.).

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. mustulemun*	AD_4	G-must
1	*G. darwinii*	AD_5	G-darw
12	*G. herbaceum*	A_1	$A_{1\text{-}8\text{-}1}$, $A_{1\text{-}8\text{-}2}$, $A_{1\text{-}5}$, $A_{1\text{-}9}$, $A_{1\text{-}17}$, $A_{1\text{-}18}$, $A_{1\text{-}19}$, $A_{1\text{-}22}$, $A_{1\text{-}23}$, $A_{1\text{-}40}$, $A_{1\text{-}49}$, and $A_{1\text{-}52}$
11	*G. arboreum*	A_2	$A_{2\text{-}8}$, $A_{2\text{-}41}$, $A_{2\text{-}47}$, $A_{2\text{-}61}$, $A_{2\text{-}72}$, $A_{2\text{-}82}$, $A_{2\text{-}106}$, $A_{2\text{-}141}$, $A_{2\text{-}194}$, $A_{2\text{-}234}$, and $A_{2\text{-}241}$
7	*G. thurberi*	D_1	D_1, $D_{1\text{-}4}$, $D_{1\text{-}23}$, $D_{1\text{-}24}$, $D_{1\text{-}35}$, $D_{1\text{-}37}$, and $D_{1\text{-}35XD8\text{-}6}$
5	*G. armourianum*	D_2	$D_{2\text{-}1}$, $D_{2\text{-}2}$, $D_{2\text{-}q}$, $D_{2\text{-}w}$, and $D_{2\text{-}19XD2\text{-}17}$
5	*G. davidsonii*	D_3	$D_{3\text{-}1}$, $D_{3\text{-}2}$, $D_{3\text{-}23}$, $D_{3\text{-}26}$, and $D_{3\text{-}28}$
32	*G. aridium*	D_4	$D_{4\text{-}1\text{-}P}$ (US004), $D_{4\text{-}2\text{-}P}$ (US005), $D_{4\text{-}3\text{-}O}$ (US010), $D_{4\text{-}4\text{-}O}$ (US011), $D_{4\text{-}5\text{-}O}$ (US012), $D_{4\text{-}6\text{-}O}$ (US013), $D_{4\text{-}7\text{-}O}$ (US15), $D_{4\text{-}8\text{-}O}$ (US016), $D_{4\text{-}9\text{-}O}$ (US017), $D_{4\text{-}10\text{-}O}$ (US041), $D_{4\text{-}11\text{-}G}$ (US072), $D_{4\text{-}12\text{-}G}$ (US076), $D_{4\text{-}13\text{-}G}$ (US078), $D_{4\text{-}14\text{-}G}$ (US080), $D_{4\text{-}15\text{-}G}$ (US081), $D_{4\text{-}16\text{-}C}$ (US117), $D_{4\text{-}17\text{-}C}$ (US120), $D_{4\text{-}18\text{-}C}$ (US121), $D_{4\text{-}19\text{-}C}$ (US122), $D_{4\text{-}20\text{-}C}$ (US126), $D_{4\text{-}21\text{-}J}$ (US128), $D_{4\text{-}22\text{-}J}$ (US130), $D_{4\text{-}23\text{-}J}$ (US136), $D_{4\text{-}24\text{-}N}$ (US138), $D_{4\text{-}25\text{-}C}$ (D4-168a), $D_{4\text{-}26\text{-}C}$ (D4-168b), $D_{4\text{-}27\text{-}C}$ (D4-168c), $D_{4\text{-}28\text{-}N}$ (US147), $D_{4\text{-}29\text{-}N}$ (US148-a), $D_{4\text{-}30\text{-}N}$ (US148-b), $D_{4\text{-}31\text{-}N}$ (US149), and $D_{4\text{-}32\text{-}N}$ (US150)
3	*G. raimondii*	D_5	$D_{5\text{-}1}$, $D_{5\text{-}2}$, and $D_{5\text{-}3}$
2	*G. gossypioides*	D_6	$D_{6\text{-}1\text{-}O}$ (US043) and $D_{6\text{-}2\text{-}O}$ (US046)
10	*G. lobatum*	D_7	$D_{7\text{-}1\text{-}M}$ (US086), $D_{7\text{-}2\text{-}M}$ (US101), $D_{7\text{-}3\text{-}M}$ (US103), $D_{7\text{-}4\text{-}M}$ (US104), $D_{7\text{-}5\text{-}M}$ (US105), $D_{7\text{-}6\text{-}M}$ (US106), $D_{7\text{-}7\text{-}M}$ (US109), $D_{7\text{-}8\text{-}M}$ (US110), $D_{7\text{-}9\text{-}M}$ (US111), and $D_{7\text{-}10\text{-}M}$ (US112)
9	*G. trilobum*	D_8	$D_{8\text{-}1\text{-}M}$ (US160), $D_{8\text{-}2\text{-}M}$ (US162), $D_{8\text{-}3\text{-}M}$ (US163), $D_{8\text{-}A}$, $D_{8\text{-}B}$, $D_{8\text{-}1}$, $D_{8\text{-}6}$, $D_{8\text{-}10}$, and $D_{8\text{-}6XD1\text{-}35}$
5	*G. laxum*	D_9	$D_{9\text{-}1\text{-}G}$ (US065), $D_{9\text{-}2\text{-}G}$ (US066), $D_{9\text{-}4\text{-}G}$ (US068), $D_{9\text{-}5\text{-}G}$ (US070), and $D_{9\text{-}6\text{-}M}$ (US098)
1	*G. turneri*	D_{10}	$D_{10\text{-}1\text{-}S}$ (US156)
3	*G. shwendinammii*	D_{11}	$D_{11\text{-}1\text{-}M}$ (US083), $D_{11\text{-}2M}$ (US084), and $D_{11\text{-}3M}$ (US100)

Note: C, Colima; G, Guerrero; J, Jalisco; N, Nayarit; M, Michoacan; O, Oaxaca; P, Puebla ecotypes; S, Sonora.

widely distributed of the arborescent Mexican diploid *Gossypium* species. With random amplified polymorphic DNA and amplified fragment length polymorphism fragments, Feng et al. (2011) found that two-thirds of the loci among the *G. aridum* accessions had allelic frequencies lower than 80%. The genetic distance (GD) between *G. gossypioides* (subsection *Selera*) and species of subsection *Erioxylum* ranged between 0.64 and 0.84, and the GD between two recognized species, *G. lobatum* and *G. schwendimanii*, within subsection *Erioxylum* was 0.32. Most molecular data support the traditional classification of *Gossypium* species and the geographical ecotypes of the *G. aridum* accessions. However, these results (Feng et al. 2011) also indicated that an additional study to establish a defensible taxonomic treatment of the various taxa was still needed. A comprehensive phylogenetic analysis of the New World cottons at the DNA level is still missing. In this study, we examined 111 accessions with microsatellite or SSR markers, which included the five allotetraploids (AD genomes), 23 Old World Asiatic diploids (A genomes), and 82 New World diploid (D genomes) accessions of the subgenus *Houzingenia* from various parts of Mexico (Ulloa et al. 2006; M. Ulloa and J.M. Stewart, unpublished data) to provide additional insight into phylogenetic relationships of species, population structure, and evolution of the New World cottons.

Material and methods

Plant material

One-hundred eleven cotton accessions that included five allotetraploids (AD_1–AD_5), 23 Asiatic diploids of the Old World (A_1 and A_2), and 82 diploids of the subgenus *Houzingenia* (D_1–D_{11} genomes) (Table 1) were assessed with SSR markers (see Supplementary Table S1[1]). The taxa were examined for identification codes, species names, and designation based on the currently accepted classification scheme of Fryxell (1992). Fifty-six of the accessions and (or) species (United States numbers) were newly collected in 2002–2006 (Table 1). The US designation was assigned to each accession using each collector's first letter of his last name, Ulloa and Stewart (Ulloa et al. 2006; Supplementary Table S2). The remaining accessions were from the United States Department of Agriculture – Agricultural Research Service (USDA-ARS) Cotton Germplasm Collection at College Station, Texas, USA. These accessions were from the northwestern part of the range of *G. turneri* collected in the Mexican state of Sonora to the southern part of the range of *G. aridum* collected in the Mexican state of Oaxaca. Because the main objective of the work was not to detect evidence of introgression within accessions, but rather to obtain a broad measure of diversity within the subgenus, except for *G. aridum* collected ecotypes, only one plant per accession was examined. However, each plant from each accession represented an in situ habitat of more than three shrubs or trees. Descriptions of the main collections that were the source of the studied accessions are available in Supplementary Table S2 (also at http//www.lbk.ars.usda.govpsgd/index-cotton.aspx/).

Marker analysis

The protocols of DNA extraction and amplification of microsatellite or SSR markers and their resolution on agarose and polyacrylamide gels were performed according to Park et al. (2005) and Frelichowski et al. (2006). We used BNL, MUCS, MUSB, and MUSS SSR markers from the 26 cotton chromosomes. From the 124 SSR used in this study, 76 SSR are known to be located on the chromosomes of the At subgenome and 48 on the chromosomes of the Dt subgenome of cotton (see Supplementary Table S1) (Yu et al. 2010, 2011; CMD: www.cottonmarker.org). PCR amplification of these cotton molecular markers was performed in a total volume of 15 μL containing 20 ng of template genomic DNA, 0.1 μmol/L of each primer (forward and reverse), 1× PCR buffer, 0.2 mmol/L dNTPs, and 1 U of *Taq* polymerase (Amplitaq, Applied Biosystems, Foster City, California, USA) with the following cycling profile: 1 cycle of 2 min at 94 °C, 10 cycles of 15 s at

[1]Supplementary data are available with the article through the journal Web site (http://nrcresearchpress.com/doi/suppl/10.1139/cjb-2012-0192).

Table 2. Analysis of molecular variance (AMOVA) estimated using ARLEQUIN 2.0. (Schneider et al. 2000) for the 111 that included 18 different genomes (AD, A, and D) of cotton (*Gossypium* spp.), including the New Word cottons.

Source of variation	df	Sum of squares	Variance components	Percentage of variation	*P* value
All-genome accessions panel[a]					
Among populations	13	4061.05	38.17	55.87	<0.0001
Within populations	97	2925.25	30.16	44.13	<0.0001
Total	110	6986.30	68.33		
Only D-genome accessions panel					
Among populations	9	2037.4	28.14	46.41	<0.0001
Within populations	69	2242.1	32.50	53.59	<0.0001
Total	78	4279.5	60.64		
Only D-genome accessions panel with Q groups at $K = 5$					
Among populations	5	1506.0	20.62	34.95	<0.0001
Within populations	76	2917.3	38.38	65.05	<0.0001
Total	81	4423.3	59.00		

[a]Panel $AD_{1\text{-}5}$ genomes, A_1 genome, A_2 genome, and D_1–D_{11} genome accessions.

94 °C, 30 s at 60 °C (step –0.5 °C/cycle for cycles 2–10), and 1 min at 72 °C, 35 cycles of 15 s at 94 °C, 30 s at 55 °C, and 1 min at 72 °C.

PCR products were separated on a 3% super-fine resolution (SFR™) agarose gel (Amresco, Solon, Ohio, USA) containing 1× TBE at 95 volts for 4–5 h, and visualized by Alphaimager software (version 5.5, Alpha Innotech Corporation, San Leandro, California, USA) after staining with ethidium bromide. Primer pairs resulting in discrete PCR banding patterns (DNA fragments) denoting a molecular marker were scored.

Molecular genetic diversity and phylogenetic analyses

The numerical taxonomy and multivariate analysis system (NTSYS-pc) version 2.1 (Rohlf 2002) and the phylogenetic analysis using parsimony (PAUP* 4.0 beta) (Swofford 2002) were used to calculate the GDs (Lynch 1990; Nei 1972) and to analyze phylogenetic relationships of *Gossypium* species by unweighted pair group method mean (UPGMA) clustering analysis. Different methods (parsimony, neighbor-joining (N-J), and UPGMA (Saitou and Nei 1987)) and 1000× bootstrapping (Felsenstein 1985) were used to statistically support the groups using PAUP software (Swofford 2002). Trees were presented using the tree-figure drawing tool FigTree version 1.31. (http://tree.bio.ed.ac.uk/) from the output file of PAUP (arbitrarily rooted and unrooted N-J tree were chosen). Only PCR products or DNA fragments that were observed with good amplification were used for data analysis.

Genetic variation within and among predefined groups and pairwise F_{ST} GDs were measured by analysis of molecular variance (AMOVA) (Reynolds et al. 1983; Weir and Cockerham 1984; Excoffier et al. 1992) using ARLEQUIN 2.0 (Schneider et al. 2000). The polymorphic information content was analyzed using the PowerMarker software package (Liu and Muse 2005). In all genetic analyses, the SSR data sets were filtered for a 5% minor or low frequency of alleles to avoid ascertainment bias. Inference of population structure (a model-based approach) implemented in the software package Structure (Pritchard et al. 2000; Pritchard and Wen 2004) for dominant markers (coded as 1–9; 2–9), was used to identify subgroups of cotton accessions. In this type of analysis, a Bayesian model-based clustering algorithm is implemented, and a predetermined number of populations (*K*) are allocated to genotyped accessions. Individuals are also allowed to be products of admixture between two or more populations. The proportion of its genome is derived from each population (q_k). We used admixture co-ancestry model under correlated allele frequencies using the burn-in time of 50 000 and the number of replications at 100 000 (Pritchard and Wen 2004) with the *K* up to 10. The statistical analysis of Structure test results (through measurements of ad hoc (ΔK) quantity of Evanno statistics (Evanno et al. 2005)) helped us to identify the real number of *K* populations for our germplasm accessions.

Results

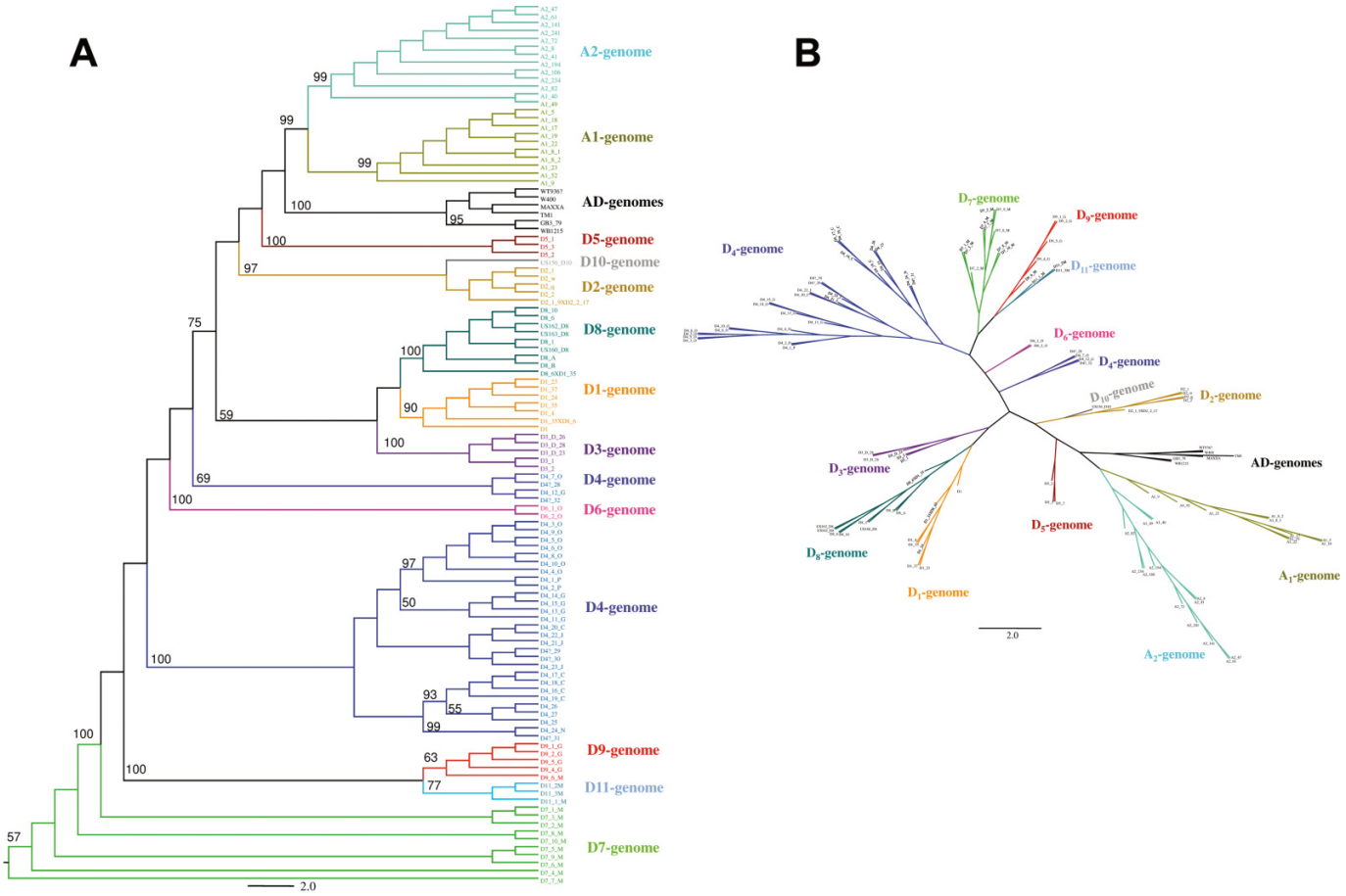
Genetic diversity and phylogenetic analyses by SSR markers

One-hundred twenty-four primer pairs amplified 694 polymorphic SSR amplicons or alleles with a mean of six alleles per SSR marker primer pair (range from 2 to 13 alleles) in all 111 accessions representing the five allotetraploids (AD_1–AD_5), the 23 Asiatic diploid species of the Old World (A_1 and A_2), and the 82 accessions of the subgenus *Houzingenia* (D_1–D_{11} genomes) of the species of the New World (Table 1). From the 694 alleles, 256 (37%) were minor or low frequency alleles and observed in only 5% of the genotyped accessions. The remaining 438 (63%) SSR alleles were highly polymorphic. When only the 82 diploid-accessions of the New World representing the subgenus *Houzingenia* (D_1–D_{11} genomes) (Table 1) were considered, SSRs yielded 560 polymorphic alleles with a mean of five alleles per primer pair. From the 560 alleles, 198 (35%) were minor or low frequency alleles and observed in only 5% of the genotyped accessions. The remaining 362 (65%) alleles were highly polymorphic. The overall polymorphic information content for the highly polymorphic SSRs ranged from 0.10 to 0.49 with a mean of 0.24.

To estimate genetic diversity within and among genomic groups, we performed the Wright's F_{ST} index using AMOVA test. In the analyses of the data sets for all genomes and the diploid New World D-genome accessions, the differentiation among groups was highly significant ($P \leq 0.0001$). A great deal of total genetic variance was attributed to the difference among and within groups (Table 2).

The highly polymorphic SSR alleles (438) were further investigated to assess the genetic diversity among and between the 111 accessions from all genomes. The mean GDs between the two most important New World tetraploid, Upland TM-1 (AD_1) and Pima 3-79 (AD_2) cottons was 0.39. Between tetraploid AD_4 genome and A_1 ancestral genome, GD mean was 0.26, and between AD_1 (Upland TM-1) and A_2 ancestral genome, GD mean was 0.38. The GD ranged between 0.19 and 0.41 in the comparisons of the shrub type sections (*Houzingenia* (D_1 and D_8), *Integrifolia* ($D_{3\text{-}d}$ and $D_{3\text{-}k}$), *Caducibracteolata* ($D_{2\text{-}1}$, $D_{2\text{-}2}$, and D_{10})) and arborescent type sections (*Erioxylum* (US-72, D_4, D_7, D_9, and D_{11}), *Selera* (D_6), and *Austroamericana* (D_5)). The GD of the arborescent *Gossypium* species ranged between 0.19 and 0.40. The two *G. gossypioides* accessions (subsection *Selera*) were quite similar to each other with a GD of only 0.04, but they were genetically distant from species of the subsection *Erioxylum* (ranged between 0.22 and 0.37). The new taxon ($D_{4\text{-}11\text{-}G}$, US-072) was genetically distant from all species examined. The smallest GD (0.22) was between US-072 and *G. laxum* ($D_{9\text{-}2\text{-}G}$). The 0.22 GD value was smaller than with any other arborescent species. *Gossypium schwendimanii* was genetically closer to *G. laxum* (0.21–0.27) than to other taxa. However, $D_{11\text{-}1\text{-}M}$ genome accession

Fig. 1. (A) Phylogenetic neighbor-joining dendrogram arbitrarily rooted of *Gossypium* accessions based on the proportion of alleles between accessions and according to the two clusters (Q1 and Q2) identified by the program Structure (Pritchard et al. 2000). (B) Unrooted neighbor-joining dendrogram of *Gossypium* accessions. Branch lengths are shown. Groups of cotton species and accessions specific to each ecotype are color-coded (online version only) for simplicity.



originally identified (Ulloa et al. 2006) as *G. schwendimanii* had the lowest GD (0.19) with $D_{4\text{-}3\text{-}O}$, $D_{4\text{-}4\text{-}O}$, $D_{4\text{-}5\text{-}O}$, and $D_{4\text{-}9\text{-}O}$ of the *G. aridum* species (accessions presented in Table 1).

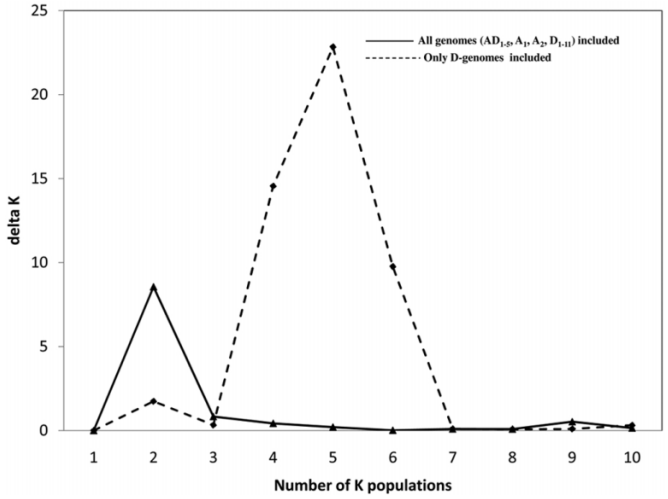
The GD among accessions of *G. aridum*, as currently circumscribed, were greater than within any other group (mean GD = 0.25 between collected accessions). These accessions were collected from seven Mexican states and clustered into distinct ecotypes that, for the most part, reflected their origin (Table 1; Figs. 1A and 1B). While GD among the Oaxaca accessions was low (GD averaging 0.15), this group was distinct from the other accessions of *G. aridum*. Each ecotype group (Colima, Guerrero, Jalisco, Nayarit, Michoacan, Oaxaca, and Puebla) distinguished between each other by the GD, ranging around 0.20–0.41. Molecular diversity reflected the differences in habitats (altitude, rainfall, topography, and geographical distribution) of the various accessions (Figs. 1A and 1B; Supplementary Table S2). Based on GD, in addition to US-072 new taxon, five new collected accessions ($D_{4\text{-}10\text{-}O}$ (US-041) GD ranging from 0.21 to 0.31, $D_{4\text{-}2\text{-}P}$ (US-005) GD ranging from 0.22 to 0.36, $D_{4\text{-}12\text{-}G}$ (US-076) GD ranging from 0.20 to 0.41, $D_{4\text{-}19\text{-}C}$ (US-122) GD ranging from 0.26 to 0.41, and $D_{4\text{-}32\text{-}N}$ (US-150) GD ranging from 0.22 to 0.36) from five different ecotypes had the larger GD against any other *G. aridum* accession from a different ecotype and any other recognized species.

Phylogenetic analysis clearly grouped all genomic groups into distinct phylogenetic clusters consistent with genomic origin, evolutionary history, and geographic distribution or ecotypes of these accessions (Figs. 1A and 1B). This phylogenetic clustering also indicated the existence of clear structured population groups, suggesting a need for a detailed analysis of population structure (Fig. 1B). Results from these analyses may also suggest that this SSR marker set could be used in other collections and other species in the genus to distinguish phylogenetic relationships.

Assessment of population structure and kinship

To assess the population structure of the 111 germplasm accessions (Table 1; Figs. 1A and 1B), a model-based Structure analysis (Pritchard et al. 2000; Pritchard and Wen 2004) was used with the population groups or clusters (K) to generate Q matrix varying from $K = 1$ up to 10 with 10 replicate-runs of each independent K. The statistical analysis of Structure (Evanno et al. 2005) output identified the maximal ΔK of the real number of K populations (Fig. 2) of our germplasm accessions. The highest value of ΔK for the cotton accessions representing all cotton genomes (A_1, A_2, D_1–D_{11}, and AD_1–AD_5 genomes) occurred at $K = 2$ (Fig. 2), indicating that these accessions consisted of two real K populations. Cotton accessions of population structure plots at $K = 2$ grouped by the Q matrix (Fig. 3C) included the Old World Asiatic diploids (A_1 and A_2) as the first group (Q1), whereas the second group (Q2) included all of the New World D-genome accessions (D_1–D_{11}). The AD-genome (AD_1–AD_5) tetraploid cottons shared alleles from both Q1 and Q2 groups. In addition, these results were mostly in agreement with N-J analysis (Figs. 1A and 1B) that showed clear groupings of A- and D-genome cotton accessions and in between the AD-genome cottons. Moreover, pairwise F_{ST} distances placed AD genomes about half way between A and D genomes (Table 2). Collectively the

Fig. 2. Ad hoc quantity (delta *K*) estimate to determine real number of *K* populations from the Structure analysis of the cotton accessions.



results from this study support the fundamental knowledge of tetraploid AD-genome formation, which was a result of the amalgamation of A and D genomes of the Old World and the New World ancestral species, respectively (Fig. 3C).

A separate analysis of population structure of only 82 diploid New World D-genome (D_1–D_{11}) accessions revealed several distinct populations within these D-genome species (Figs. 2 and 3). The highest value of Δ*K* for the cotton accessions representing the D genome occurred at *K* = 5 (Fig. 2), indicating that the D-genome accessions consisted of five real *K* populations. Cotton accessions of population structure plots at *K* = 5 grouped by the Q matrix (Figs. 3A and 3B) included the D_3 accessions as a Q1 cluster with five accessions (Table 1); the D_1, D_2, and D_5 accessions as a Q2 cluster with 15 accessions from the following genomes: the D_8 accessions as a Q3 cluster with nine accessions (Table 1; Figs. 3A and 3B); the D_4 and D_{11} accessions as a Q4 cluster with 30 accessions from these two genomes; and the D_7 accessions as a Q5 cluster with 10 accessions (magneta) (Table 1). There were 13 D-genome accessions (D_4, D_6, D_9, D_{10}, and D_{11}) that shared more or less genetic structure with one of the above clusters and placed as mixed samples (Fig. 3B; Table 4). AMOVA and F_{ST} estimates using these Q clusters or groups suggested that the differentiation among groups was highly significant ($P \leq 0.0001$) (Table 3). We observed that 34.95% of the total genetic variance was attributed to the difference among Q cluster or groups, and the remaining variance (65.05%) was attributed to within groups (Table 2). A great deal of genetic differentiation (>50%) occurred between Q1 (D_3 genome) and Q3 (D_8 genome), Q1 and Q5 (D_7 genome), and between Q3 and Q5 (Table 4).

Discussion

Knowledge of the genetic diversity and population structure is one of the first steps to understanding the preservation and the utilization of genetic resources for crop improvement. To assess this genetic diversity and population structure in *Gossypium* species of the New World, 111 cotton accessions representing five allotetraploids (AD_1–AD_5 genomes), 23 Old World Asiatic diploids (A_1 and A_2 genomes), and 82 New World diploid species (included yet unsubscribed taxa of the subgenus *Houzingenia* (D_1–D_{11} genomes)) were assessed with SSR markers with wide genome coverage. Based on highly polymorphic SSR markers, genetic diversity, phylogeny, and population structure analyses supported the genomic origin, evolutionary history, and geographic distribution of these accessions (or ectotypes) (Table 1; Figs. 1–3). Assessments of population structure (highest Δ*K* occurred at *K* = 2) and phylogenetic analyses yielded consistent results with the fundamental knowledge of the origin of allopolyploids, which was a result of the amalgamation of A and D genomes of the Old World and the New World ancestral species, respectively. Interestingly, a separate analysis of population structure (highest Δ*K* occurred at *K* = 5) of the New World D genome (D_1–D_{11}) cottons revealed that isolation and distance of habitats appear to be the main factor of greater GD between two different species accessions of the D genome.

Our SSR data provided useful insight information into the geographical, taxonomical distribution, and evolutionary history of the New World cottons. When considering the GD and structured groups (Q1–Q5), SSRs supported that interspecific and molecular differentiation of these species were more common that appreciated in angiosperm evolution (Álvarez et al. 2005). When AFLP fragments were used to compare between nonsympatric arborescent species, unexpected results were observed (Feng et al. 2011). Similar unexpected results were observed using SSR data in our study. *Gossypium lobatum* and *G. schwendimanii* are two morphologically distinct species (Fryxell 1992) and are sympatric in part of their range. These two species are thought to hybridize in their native habitats (Ulloa et al. 2006). However, *G. schwendimanii* was genetically closer to *G. laxum* (GD = 0.21–0.27) a nonsympatric species instead of *G. lobatum* (sympatric, GD = 0.23–0.31). When Álvarez and Wendel (2006) and Feng et al. (2011) compared the new taxon ($D_{4\text{-}11\text{-}G}$, US-072) with *G. laxum* ($D_{9\text{-}2\text{-}G}$) using the AFLP marker system, similar results were reported of a nonsympatric versus sympatric species comparison. Our SSR marker system also revealed similar GD (0.22) estimates between ($D_{4\text{-}11\text{-}G}$, US-072) and *G. laxum* ($D_{9\text{-}2\text{-}G}$) to the above mentioned studies.

Observations of *G. aridum* accessions indicated extensive differences in leaf size, vesture of leaves, morphology in the lysigenous glands on the capsule, and period of flowering. These morphological differences reflected for the most part the accessions habitat or best described their ecotype, especially period of flowering (M. Ulloa and J.M. Stewart, personal observations and comparison among herbarium specimens in MEXU). Based on GD, in addition to US-072 new taxon, five new collected accessions ($D_{4\text{-}10\text{-}O}$ (US-041), $D_{4\text{-}2\text{-}P}$ (US-005), $D_{4\text{-}12\text{-}G}$ (US-076), $D_{4\text{-}19\text{-}C}$ (US-122), and $D_{4\text{-}32\text{-}N}$ (US-150)) from five different ecotypes had the larger GD when compared with any other *G. aridum* accession and any other recognized *Gossypium* species of the D genome, GD > 0.19 and GD ≤ 0.41, respectively. Feng et al. (2011) reported that the GD between the *G. aridum* accessions from Colima and accessions from other Mexican states ranged from 0.32 to 0.38, which was equal to or greater than the GD between two recognized D-genome species. In this study, similar GD mean (0.28) was detected between any of the above five newly collected accessions of the *G. aridum* with any other recognized *Gossypium* species of the D genome or any other different ecotypes of *G. aridum*. Based in our GD comparison of the diploid D genome of the New World cottons, we propose a minimum mean threshold of GD = 0.20 for distinguishing a new taxon, pending morphological taxonomy. This threshold will definitely distinguish subspecies and (or) ecotypes from different habitats. The five newly collected accessions reported herein were also genetically distant between the different ecotypes and from all species examined, and may be proposed as new taxa.

Our phylogenetic analysis grouped all species into distinct phylogenetic groups consistent with genomic origin and evolutionary history. In addition, our analysis clustered the diploids of the New World into six sections with the three bushy types (*Houzingenia* (D_1 and D_8), *Integrifolia* ($D_{3\text{-}d}$), and *Caducibracteolata* ($D_{2\text{-}1}$, $D_{2\text{-}2}$, and D_{10})) and three arborescent types (*Erioxylum* (US-072, D_4, D_7, D_9, and D_{11}), *Selera* (D_6), and *Austroamericana* (D_5)). The circumscriptions of the subgenus and species boundaries are well understood (Fryxell 1979, 1992). Our results are in agreement with other studies (Cronn et al. 1999; Small and Wendel 2000*b*; Álvarez et al. 2005) (Figs. 1A and 1B).

Fig. 3. Population structure of cotton (*Gossypium* spp.) accessions: (A) D-genome cotton accessions population structure plots at $K = 5$ grouped by Q matrix, (B) grouped by genomic origins of the D-genome accessions, and (C) all-genome cotton accessions population structure plot at $K = 2$, where Q1 (red) represents diploid A-genome accessions, Q2 (green) represent diploid Dgenomes, and red and green represents the amalgamation of the tetraploid AD genomes. Colour coding is for the online version only.

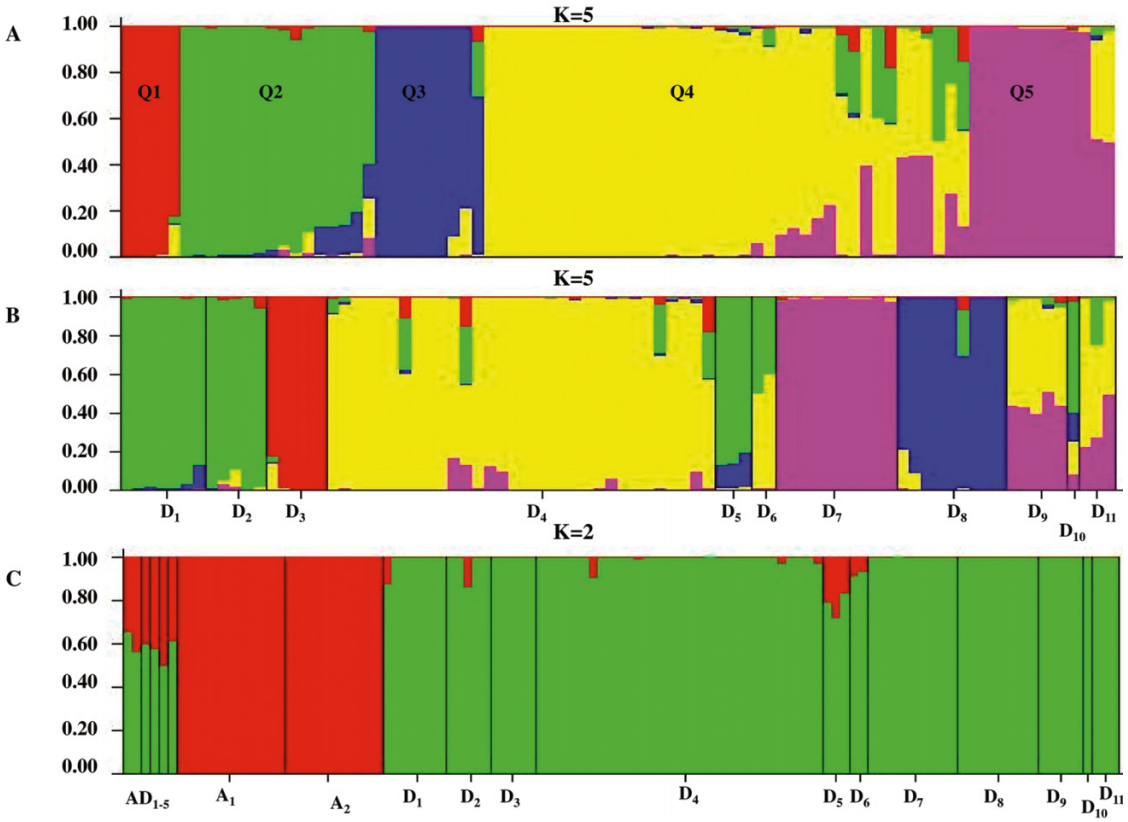


Table 3. Pairwise and population-specific F_{ST} of accessions representing all cotton genomes in this study.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	AD genomes	0.51													
2	A_1 genome	0.53	0.59												
3	A_2 genome	0.60	0.61	0.60											
4	D_5 genome	0.41	0.77	0.83	0.60										
5	D_1 genome	0.55	0.77	0.79	0.69	0.57									
6	D_2 genome	0.42	0.73	0.75	0.53	0.52	0.55								
7	D_3 genome	0.54	0.80	0.84	0.76	0.62	0.56	0.59							
8	D_4 genome	0.41	0.54	0.56	0.40	0.39	0.32	0.38	0.52						
9	D_6 genome	0.45	0.80*	0.85	0.82*	0.64	0.55	0.76*	0.28	0.61					
10	D_7 genome	0.54	0.75	0.77	0.63	0.59	0.51	0.61	0.29	0.60	0.56				
11	D_8 genome	0.65	0.83	0.85	0.78	0.65	0.68	0.76	0.42	0.80	0.67	0.59			
12	D_{10} genome	0.29*	0.80*	0.85*	0.76*	0.57*	0.35*	0.73*	0.20*	0.88*	0.53*	0.76*	0.62*		
13	D_9 genome	0.45	0.74	0.78	0.60	0.56	0.47	0.60	0.23	0.53	0.41	0.67	0.45*	0.56	
14	D_{11} genome	0.39	0.74	0.79	0.60*	0.53	0.41	0.60	0.13	0.57*	0.37	0.68	0.36*	0.24	0.57

Note: Diagonal elements are population-specific F_{ST}. Below-diagonal elements are pairwise F_{ST}.
*Nonsignificant at $P < 0.05$, remaining comparisons are significant at $P < 0.05$.

Table 4. Pairwise and population-specific F_{ST} in a five groups of D-genome accessions derived from Structure analysis.

Populations	1	2	3	4	5	6
Q1_D3_5 samples	0.38					
Q2_D1_D2_D5_15 samples	0.36	0.34				
Q3_D8_9 samples	0.77	0.42	0.38			
Q4_D4_D11_30 samples	0.43	0.27	0.45	0.34		
Q5_D7_10 samples	0.63	0.37	0.67	0.31	0.37	
Mixed_D4_D6_D9_D10_D11_13 samples	0.38	0.19	0.45	0.14	0.28	0.34

Note: Diagonal elements are population-specific F_{ST}. Below-diagonal elements are pairwise F_{ST}.
*All comparisons are significant at $P < 0.0001$.

The history of cotton started with the evolution of the genus around 10–20 million years ago (Wendel and Albert 1992; Percival et al. 1999), starting with the formation or origin of the American diploids or New World cottons that may be estimated at around 6.7 million years ago (Wendel and Albert 1992). Following that was the allopolyploid formation around 1.5 million years ago (Senchina et al. 2003). Population structure and population-differentiation analyses support the fundamental fact of the amalgamation and evolutionary history of the A and D genomes to form the allotetraploid (AD) cottons (Fig. 3C). The existence of confounding population structure has been reported from analyses of many plant populations (Pritchard et al. 2000; Zhao et al. 2007) and germplasm resources (Abdurakhmonov and Abdukarimov 2008).

Population structure also shed light on the emergence and dispersion of the diploids of the New World. Similar consistency with the New World cottons (Cronn et al. 1999; Small and Wendel 2000*b*; Álvarez et al. 2005) was observed in the D-genome population strata, in which a rapid radiation of the American diploid cotton linage took place somewhere in southwestern Mexico, and it was followed by a differentiation-speciation (Small and Wendel 2000*b*). It has been suggested that this radiation might have occurred before the separation of the Baja California peninsula (7–12 million years ago) from mainland Mexico (Álvarez et al. 2005). Our population structure analyses 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). These two species are clearly distinguished on many morphological features, as follows: leaves, flowers, seed capsule, pubescence, etc. This colonization event of the species in the Baja California peninsula has previously been suggested by phylogenetic analyses (Álvarez and Wendel 2006).

A coalescent group (Q2) composed of subsections *Houzingenia* (D_1 accessions), *Caducibracteata* ($D_{2\text{-}1}$ accessions), and *Austroamericana* (D_5 accessions) seems to share alleles from a common ancestor (Figs. 3A and 3B). In data supported by phylogenic studies, subsections *Houzingenia* and *Austroamericana* have been reported to form monophylic groups (Álvarez et al. 2005). Based on geographic distribution and morphology of some of these species, in conjunction with the population structure analyses, we can hypothesize that the New World diploid species may derive from five major lineages (Q1–Q5) that eventually radiate and differentiate about 7–8 million years ago through the country of Mexico. Species like *G. gossypioides* (D_6 genome), *G. laxum* (D_9 genome), *G. turneri* (D_{10} genome), and *G. shwendinammii* (D_{11} genome) experienced a more recent differentiation event (Fig. 4B). Possibly, supra-specific coalescence of some alleles in these species may support the mixed sample group of the 13 D-genome accessions (D_4, D_6, D_9, D_{10}, and D_{11}), experiencing more recent hybridization events.

The arborescent subsection *Erioxylum* are among the most distinctive in the genus (Fryxell 1992). However, the sectional levels of *G. aridum*, as currently circumscribed (Fryxell 1979; Ulloa et al. 2006), remain unresolved. SSR markers have proved to be a powerful tool in elucidating genetic relationships and population structure of these accessions. The proposed GD minimum threshold (0.20) in this study may be useful to define a new taxon, and unclear relationships among species or genetically distant geographical accession ecotypes of *G. aridum* may be resolved. In addition, GD threshold will help to place newly collected accessions in corresponding species groups. Continued efforts of collection, evaluations, and utilization of a broader range of cotton germplasm, including diploid species of the *Gossypium* genus, will increase the genetic diversity of cultivated cottons.

Acknowledgments

In memory of James McD. Stewart. The authors would like to offer their appreciation and heartfelt thanks to the many individuals, too numerous to name, residing in the various states who contributed invaluable service, guidance, and help, without which the collection of some of these accessions would have been much less fruitful. Special thanks to Young Hoon Park, a postdoctoral fellow at the USDA-ARS, WICRU, Shafter, California, for helping with electrophoresis and fingerprinting of these accessions during 2004–2005. This study was partially supported by a specific cooperative agreement between USDA-ARS and the Mexican agency INIFAP (ARIS Log Nos. 5303-21220-001-10S and 5303-2-F159). Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA. The USDA is an equal opportunity provider and employer.

References

Abdalla, A.M., Raddy, O.U.K., El-Zik, K.M., and Pepper, A.E. 2001. Genetic diversity and relationships of diploid and tetraploid cottons revealed using AFLP. Theor. Appl. Genet. **102**: 222–229. doi:10.1007/s001220051639.

Abdurakhmonov, I.Y. 2007. Exploiting genetic diversity. Proceedings of World Cotton Research Conference-4, P2153, Lubbock, Texas, U.S.A., September 10–14, 2007.

Abdurakhmonov, I.Y., and Abdukarimov, A. 2008. Application of association mapping to understanding the genetic diversity of plant germplasm resources. International Journal of Plant Genomics, **574927**: 1–18. [ISSN 1687-5389.]

Abdurakhmonov, I.Y., Buriev, Z.T., Shermatov, S.E., Abdullaev, A., Urmonov, K., Kushanov, F., Egamberdiev, S.S., Shapulatov, U., Abdukarimov, A., Saha, S., Jenkins, J.N., Kohel, R.J., Yu, J.Z., Pepper, A.E., Kumpatla, S.P., and Ulloa, M. 2011. Genetic diversity in Gossypium genus. In Genetic diversity book edited by Mahmut ÇALIŞKAN. Intech Publications. [In press.]

Álvarez, I., and Wendel, J.F. 2006. Ancient interspecific introgression and modern genetic differentiation within *Gossypium aridum* (Malvaceae) and its relatives. Evolution, **60**: 505–517. doi:10.1554/05-184.1. PMID:16637496.

Álvarez, I., Cronn, R., and Wendel, J.F. 2005. Phylogeny of the New World diploid cottons (*Gossypium* L., Malvaceae) based on sequences of three low-copy nuclear genes. Plant Syst. Evol. **252**: 199–214. doi:10.1007/s00606-004-0294-0.

Buckler, E.S., Ippolito, A., and Holtsford, T.P. 1997. The evolution of ribosomal DNA: divergent paralogues and phylogenetic implications. Genetics, **145**: 821–832. PMID:9055091.

Cronn, R.C., Small, R.L., and Wendel, J.F. 1999. Duplicated genes evolve independently after polyploid formation in cotton. Proc. Natl. Acad. Sci. **96**: 14406–14411. doi:10.1073/pnas.96.25.14406. PMID:10588718.

Evanno, G., Regnaut, S., and Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. **14**(8): 2611–2620. doi:10.1111/j.1365-294X.2005.02553.x. PMID:15969739.

Excoffier, L., Smouse, P., and Quattro, J. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics, **131**: 479–491. PMID: 1644282.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution, **39**: 783–791. doi:10.2307/2408678.

Feng, C., Ulloa, M., Perez-M., C., Stewart, J., McD . 2011. Distribution and molecular diversity of arborescent *Gossypium* species. Botany, **89**(9): 615–624. doi: 10.1139/b11-042.

Frelichowski, J.E., Palmer, M.B., Main, D., Tomkins, J.P., Cantrell, R.G., Stelly, D.M., Yu, J., Kohel, R.J., and Ulloa, M. 2006. Cotton genome mapping with new microsatellites from Acala 'Maxxa' BAC-ends. Mol. Genet. Genomics, **275**: 479–491.

Fryxell, P.A. 1979. The natural history of the cotton tribe. Texas A&M University Press, College Station, Tex.

Fryxell, P.A. 1992. A revised taxonomic interpretation of *Gossypium* L. (Malvaceae). Rheedea, **2**: 108–165.

Fryxell, P.A., Craven, L.A., and Stewart, J.M. 1992. A revision of *Gossypium* sect. *Grandicalyx* (Malvaceae), including the description of six new species. Syst. Bot. **17**: 91–114.

Hanson, R.E., Zhao, X.P., Islam-Faridi, M.N., Paterson, A.H., Zwick, M.S., Crane, C.F., McKnight, T.D., Stelly, D.M., and Price, H.J. 1998. Evolution of interspersed repetitive elements in Gossypium (Malvaceae). Am. J. Bot. **85**: 1364–1368. doi:10.2307/2446394. PMID:21684890.

Hanson, R.E., Islam-Faridi, M.N., Crane, C.F., Zwick, M.S., Czeschin, D.G., Wendel, J.F., McKnight, T.D., Price, H.J., and Stelly, D.M. 1999. Ty1-copia-retrotransposon behavior in a polyploidy cotton. Chromosome Res. **8**: 73–76. PMID:10730591.

Kantartzi, S.K., Ulloa, M., Sacks, E., and Stewart, J.McD. 2009. Assessing genetic diversity in *Gossypium arboreum* L. cultivars using genomic and EST-derived microsatellites. Genetica, **136**: 141–147.

Linos, A.A., Bebeli, P.J., and Kaltsikes, P.J. 2002. Cultivar identification in upland cotton using RAPD markers. Aust. J. Agric. Res. **53**: 637–642. doi:10.1071/AR01129.

Liu, K., and Muse, S.V. 2005. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics, **21**: 2128–2129. doi:10.1093/bioinformatics/bti282. PMID:15705655.

Liu, Q., Brubaker, C.L., Green, A.G., Marshall, D.R., Sharp, P.J., and Singh, S.P. 2001. Evolution of the FAD2-1 fatty acid desaturase 5′ UTR intron and the molecular systematics of Gossypium (Malvaceae). Am. J. Bot. **88**(1): 92–102. doi:10.2307/2657130. PMID:11159130.

Lynch, M. 1990. The similarity index and DNA fingerprinting. Mol. Biol. Evol. **7**: 478–484. PMID:2263197.

Meredith, W.R.J. 1992. Improving fiber strength through genetics and breeding. *In* Cotton Fiber Cellulose: Structures, Functions, and Utilization Conference. *Edited by* C.R. Benedict and G.M. Jividen. National Cotton Council, Memphis, Tenn. pp. 289–302.

Multani, D.S., and Lyon, B.R. 1995. Genetic fingerprinting of Australian cotton cultivars with RAPD markers. Genome, **38**(5): 1005–1008. doi:10.1139/g95-132. PMID:18470223.

Nei, M. 1972. Genetic distance between populations. Am. Nat. **106**: 283–292. doi:10.1086/282771.

Park, Y.H., Alabady, M.S., Ulloa, M., Sickler, B., Wilkins, T.A., Yu, J., Stelly, D.M., Kohel, R.J., El-Shihy, O.M., and Cantrell, R.G. 2005. Genetic mapping of new cotton fiber loci using EST-derived microsatellites in an interspecific recombinant inbred line (RIL) cotton population. Mol. Genet. Genomics, **274**: 428–441.

Percival, A.E., Wendel, J.F., and Stewart, J.M. 1999. Taxonomy and germplasm resources. *In* Cotton; origin, history, technology, and production. *Edited by* C.W. Smith and J.T. Cothren. Wiley, New York. pp. 33–63.

Pillay, M., and Myers, G.O. 1999. Genetic diversity in cotton assessed by variation in ribosomal RNA genes and AFLP markers. Crop Sci. **39**: 1881–1886. doi:10.2135/cropsci1999.3961881x.

Pritchard, J.K., and Wen, W. 2004. Documentation for Structure software. The University of Chicago Press, Chicago.

Pritchard, J.K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multi-locus genotype data. Genetics, **155**: 945–959. PMID: 10835412.

Rahman, M., Hussain, D., and Zafar, Y. 2002. Estimation of genetic divergence among elite cotton cultivars-genotypes by DNA fingerprinting. Crop Sci. **42**: 2137–2144. doi:10.2135/cropsci2002.2137.

Ram, S.G., Thiruvengadam, V., and Vinod, K.K. 2007. Genetic diversity among cultivars, landraces and wild relatives of rice as revealed by microsatellite markers. J. Appl. Genet. **48**(4): 337–345. doi:10.1007/BF03195230. PMID: 17998590.

Reynolds, J., Weir, B.S., and Cockerham, C.C. 1983. Estimation for the coancestry coefficient: basis for a short-term genetic distance. Genetics, **105**: 7767–777.

Rohlf, F.J. 2002. NTSYS pc: Numerical taxonomy system. Ver. 2.01. Exter Publishing, Ltd., Setauket, N.Y.

Saitou, N., and Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**: 406–425. PMID:3447015.

Schneider, S., Roessll, D., and Excoffier, L. 2000. ARLEQUIN: a software for population genetics data analysis, version 2.0. Genetics and Biometry Laboratory, Department of Anthropology, Geneva, Switzerland, University of Geneva.

Senchina, D.S., Alvarez, I., Cronn, R., Liu, B., Rong, J., Noyes, R.D., Paterson, A.H., Wing, R.A., Wilkins, T.A., and Wendel, J.F. 2003. Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. Mol. Biol. Evol. **20**: 633–643. doi:10.1093/molbev/msg065. PMID:12679546.

Small, R.L., and Wendel, J.F. 2000*a*. Phylogeny, duplication, and intraspecific variation of *Adh* sequences in new world diploid cottons (*Gossypium* L., Malvaceae). Mol. Phylogenet. Evol. **16**: 73–84. doi:10.1006/mpev.1999.0750. PMID: 10877941.

Small, R.L., and Wendel, J.F. 2000*b*. Copy number lability and evolutionary dynamics of the Adh gene family in diploid and tetraploid cotton (Gossypium). Genetics, **155**: 1913–1926. PMID:10924485.

Small, R.L., Ryburn, J.A., Cronn, R.C., Seelanan, T., and Wendel, J.F. 1998. The tortoise and the hare: choosing between noncoding plastome and nuclear Adh sequences for phylogeny reconstruction in a recently diverged plant group. Am. J. Bot. **85**: 1301–1315. doi:10.2307/2446640. PMID:21685016.

Swofford, D.L. 2002. Phylogenetic analysis using parsimony (∗ and other methods). Sinauer, Sunderland, Mass.

Tanksley, S.D., and McCouch, S.R. 1997. Seed banks and molecular maps: unlocking genetic potential from the wild. Science, **277**: 1063–1066. doi:10.1126/science.277.5329.1063. PMID:9262467.

Ulloa, M. 2006. Heritability and correlation assessments for agronomic and fiber traits in an okra upland cotton population. Crop Sci. **46**: 1508–1514. doi:10.2135/cropsci2005.08-0271.

Ulloa, M., Stewart, J.McD., Garcia-C, E.A., Godoy-A, S., Gaytán-M, A., and Acosta, N, S. 2006. Cotton genetic resources in the western states of Mexico: in situ conservation status and germplasm collection for ex situ preservation. Genet. Resour. Crop Evol. **53**(4): 653–668. doi:10.1007/s10722-004-2988-0.

Ulloa, M., Brubaker, C., and Chee, P. 2007. Cotton. *In* Genome mapping & molecular breeding. Vol. 6: Technical crops. *Edited by* C. Kole. Springer, Heidelberg, Berlin, New York, Tokyo. pp. 1–49.

Ulloa, M., Percy, R., Hutmacher, R.B., and Zhang, J. 2009. The future of cotton breeding in the Western United States. J. Cot. Sci. **4**: 246–255.

Vigouroux, Y., Glaubitz, J.C., Matsuoka, Y., Goodman, M.M., Sanchez, G.J., and Doebley, J. 2008. Population structure and genetic diversity of New World maize races assessed by DNA microsatellites. Am. J. Bot. **95**(10): 1240–1253. doi:10.3732/ajb.0800097. PMID:21632329.

Wallace, T.P., Bowman, D., Campbell, B.T., Chee, P., Gutierrez, O.A., Kohel, R.J., McCarty, J., Myers, G., Percy, R., Robinson, F., Smith, W., Stelly, D.M., Stewart, J.M., Thaxton, P., Ulloa, M., and Weaver, D.B. 2009. Status of the U.S.A. cotton germplasm collection and crop vulnerability. Genet. Resour. Crop Evol. **56**: 507–532.

Weir, B.S., and Cockerham, C.C. 1984. Estimating F-statistics for the analysis of population structure. Evolution, **38**: 1358–1370. doi:10.2307/2408641.

Wendel, J.F., and Albert, V.A. 1992. Phylogenetics of the cotton genus (*Gossypium*). Character-state weighted parsimony analysis of chloroplast-DNA restriction site data and its systematic and biogeographic implications. Syst. Bot. **17**: 115–143.

Wendel, J.F., Schnabel, A., and Seelanan, T. 1995. An unusual ribosomal DNA sequence from *Gossypium* gossypioides reveals ancient, cryptic intergenomic introgression. Mol. Phylogenet. Evol. **4**: 298–313. doi:10.1006/mpev.1995.1027.

Wendel, J.F., Cornn, R.C., Johnston, J.S., and Price, H.J. 2002. Feast and famine in plant genomes. Genetica, **115**: 37–47. doi:10.1023/A:1016020030189. PMID: 12188047.

Xu, Q., Zhang, X., Nie, N., and Feng, C.D. 2002. Genetic diversity among cotton cultivars (*G. hirsutum* L.) resistant to Fusarium wilt as estimated by RAPD markers. Sci. Agric. Sin. **35**(3): 272–276.

Yu, J.Z., Kohel, R.J., Fang, D.D., Cho, J., Van Deynze, A., Ulloa, M., Hoffman, S.M., Pepper, A.E., Stelly, D.M., Jenkins, J.N., Saha, S., Kumpatla, S.P., Shah, M.R., Hugie, W.V., and Percy, R.G. 2011. A high-density SSR and SNP genetic map of the tetraploid cotton genome. Genes, Genomes, Genetics: G3 (accepted Nov 04, 2011).

Yu, Y., Yuan, D., Liang, S., Li, X., Wang, X., Lin, Z., and Zhang, X. 2010. Genome structure of cotton revealed by a genome-wide SSR genetic map constructed from a BC_1 population between *Gossypium hirsutum* and *G. barbadense*. BMC Genomics, **21**: 15.

Zhao, K., Aranzana, M.J., Kim, S., Lister, C., Shindo, C., Tang, C., et al. (2007) An *Arabidopsis* Example of Association Mapping in Structured Samples. PLoS Genet. **31**: e4.

Zhao, X., Si, Y., Hanson, R.E., Crane, C.F., Price, H.J., Stelly, D.M., Wendel, J.F., and Paterson, A.H. 1998. Dispersed repetitive DNA has spread to new genomes since polyploidy formation in cotton. Genome Res. **8**: 479–492. PMID: 9582192.