CRYPTIC INTERSPECIFIC INTROGRESSION AND GENETIC DIFFERENTIATION WITHIN GOSSYPIUM ARIDUM (MALVACEAE) AND ITS RELATIVES

INÉS ÁLVAREZ1 AND JONATHAN F. WENDEL2
1Real Jardín Botánico de Madrid, Consejo Superior de Investigaciones Científicas, Plaza de Murillo 2, E-28014 Madrid, Spain
E-mail: ines@ma-rjb.csic.es
2Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, Iowa 50011
E-mail: jfw@iastate.edu

Abstract.—Interspecific gene flow is increasingly recognized as an important evolutionary phenomenon in plants. A surprising observation is that historical introgression is often inferred between species that presently have geographic and reproductive barriers that would appear to prohibit the inferred sexual exchange. A striking example concerns Gossypium aridum (subsection Erioxylum); previous analyses have shown that populations from Colima (southwestern Mexico) have a chloroplast genome (cpDNA) similar to that of a different taxonomic subsection (Integrifolia) that presently is confined to Baja California and the Galapagos Islands, whereas other G. aridum populations share a cpDNA lineage with each other and with other species in subsection Erioxylum. To evaluate further the possibility that this cpDNA evidence reflects introgression as opposed to some other evolutionary process, as well as to explore patterns of genetic diversity and similarity in both subsections, we conducted amplified fragment-length polymorphism (AFLP) analysis using 50 populations representing all seven species in the two subsections. Genetic diversity is high in G. aridum, and is strongly correlated with geography, as are similarities among the five species in subsection Erioxylum. This subsection is genetically distant from the two species in subsection Integrifolia, whose populations are highly similar inter se. Populations of G. aridum from Colima are genetically distinct from the remainder of the species, and exhibit a comparatively high frequency of AFLP fragments that otherwise are diagnostic of the Integrifolia lineage. These data implicate intersubsectional introgression between presently allopatric and genetically isolated clades, giving rise to a morphologically cryptic, introgressant entity. Biogeographic considerations suggest that this history was initiated following migration of one or more seeds from Baja California to the Colima coast, perhaps during the Pleistocene. We suggest that cryptic and seemingly improbable interspecific introgression and molecular differentiation may be more common than appreciated in angiosperm evolution.

Key words.—Phylogeography, genetic diversity, introgression, molecular differentiation, amplified fragment-length polymorphism, Gossypium aridum.

One important insight from comparative molecular phylogenetic analyses over the last 20 years is that interspecific gene flow and hybrid speciation are common phenomena in plants (Rieseberg and Soltis 1991; Rieseberg and Wendel 1993; Rieseberg et al. 1995; Arnold 1997; Cronn et al. 2003; Cronn and Wendel 2004; Hegarty and Hiscock 2005). Although only a small fraction of plant genera have been examined using multiple, independent phylogenetic datasets, incongruence among two or more molecular datasets has shown that hybridization has been even more prevalent than indicated by morphological and cytogenetic data. Cryptic historical introgression appears to be common in some well-studied genera, such as Iris (Arnold et al. 2004), and Ipomopsis (Campbell 2004), in which hybrid and introgressant genotypes may be more fit than their parents in novel habitats. A remarkable aspect of many recent studies is that interspecific gene flow has been inferred between species that presently show strong geographic or reproductive barriers that would appear to prohibit the inferred sexual exchange. These barriers include extreme geographic allopatry as well as intrinsic barriers to mating and genetic exchange (Rieseberg and Wendel 1993; Rieseberg et al. 1995; Arnold 1997; Rieseberg 1997, 2003; Cronn et al. 2003; Cronn and Wendel 2004).

A case in point concerns the evolutionary history of subsection Erioxylum, a group of American Gossypium species centered in southern and western Mexico. Phylogenetic affinities among American diploid cottons (Gossypium L., subgenus Houzingenia Fryxell) have been studied using multiple datasets, including chloroplast DNA (cpDNA) restriction sites (DeJonge 1992; Wendel and Albert 1992), cpDNA spacer and gene sequences (Cronn et al. 2003), nuclear ribosomal DNA (nrDNA) sequences (Cronn et al. 1996; Seelanan et al. 1997), and sequences of low-copy nuclear genes (Small and Wendel 2000; Cronn et al. 2003; Álvarez et al. 2005). All of these analyses support the monophyly of the subgenus (reviewed in Wendel and Cronn 2003) and the taxonomic recognition of six subsections (Austroamericana, Caducibracteolata, Erioxylum, Houzingenia, Integrifolia, and Selera). Notwithstanding these multiple analyses, relationships among subsections remain unresolved in some cases, in part reflecting rapid cladogenesis early in the history of this subgenus (Álvarez et al. 2005), the probable occurrence of multiple episodes of historical interspecific hybridization (Cronn et al. 2003; Cronn and Wendel 2004), and perhaps insufficient population sampling.

Among the 13 species of American diploid cottons, a group of four Mexican species (subsection Erioxylum) stand out...
morphologically by virtue of their arborescent habit (trees up to about 20 m) as well as other characters. Previous work (Small and Wendel 2000; Álvarez et al. 2005) reported high levels of intraspecific variation in the four currently recognized species (G. aridum, G. laxum, G. lobatum, and G. schwendimanii) of the subsection, as well as an absence of AdhA allele coalescence within species (Small and Wendel 2000). This trans-specific polymorphism could reflect either interspecific gene flow, a lack of coalescence following speciation from a polymorphic ancestor, or both phenomena. Earlier indications that G. aridum had an unconventional evolutionary history trace back more than a decade (Wendel and Albert 1992), when a single accession of G. aridum included in a phylogenetic analysis of cpDNA restriction-site variation was shown to share its cpDNA ancestry with members of the otherwise cladistically distant subsection Integrifolia (G. davidsonii and G. klotzschianum), rather than with the other Mexican arborescent species. These two subsections are taxonomically well defined and morphologically distinct, as well as geographically allopatric. Subsection Integrifolia, from Baja California (G. davidsonii) and the Galapagos Islands (G. klotzschianum), is represented by shrubby plants with rotate yellow corollas, laciniate to dentate involucral bracts, and sparse seed pubescence. In contrast, members of subsection Erioxylum, broadly distributed in western Mexico exclusive of Baja California, are trees with funnelform pink corollas, entire involucral bracts, and dense seed pubescence. Thus, the cpDNA data were surprising, and were hypothesized to reflect cytoplasmic introgression into G. aridum from a taxon in the subsection Integrifolia lineage. Further experiments involving the same species but deeper population sampling and both cpDNA restriction-site and isozyme analyses (DeJoode 1992) confirmed the previous observation of putative cytoplasmic introgression from Integrifolia (Fig. 1), but added the twist that this introgression is evident only in accessions of G. aridum from the single Mexican state of
Colima, a small region located on the southwestern Mexican coast. All other *G. aridum* accessed samples, from Guerrero, Jalisco, Michoacan, and Sinaloa, contain “normal” plastid genomes, cladistically monophyletic with those from all other *Erioxylum* species and populations.

Thus, *G. aridum* offers a striking example of interpopulational, intraspecific polymorphism in plastid genomes, raising the question as to how and when the polymorphism originated. Additionally, it is of interest to ask whether the putative hybridization event left footprints in the nuclear genome of *G. aridum* from Colima, and, if so, how extensive nuclear introgression might be, as well as its genomic distribution and composition. To date, evidence of nuclear introgression has not been observed, either in 5S rDNA genes (Cronn et al. 1996) or coding sequences from four independent nuclear loci (Small and Wendel 2000; Álvarez et al. 2005). However, these surveys collectively explore only a tiny portion (~7 kb) of the approximately 900 Mb nuclear genome (Hendrix and Stewart 2005) of these *Gossypium* species.

To gain further insight into the possibility of nuclear introgression in Colima populations of *G. aridum* and to explore genetic similarities with *G. aridum* and sympatric species, it is necessary to sample the genome more thoroughly. With these objectives in mind we employed amplified fragment-length polymorphism (AFLP) analysis (Vos et al. 1995). This commonly used DNA fingerprinting technique (reviewed in Mueller and Wolfenbarger 1999) has applicability to a wide array of evolutionary questions, ranging from studies of hybridization and allopolyploidy (Liu et al. 2001; Adams et al. 2004) to paternity analysis (Pertl et al. 2002; Chauhan et al. 2004) to prehistoric questions of species circumscription and delimitation (e.g., Van Den Berg et al. 2002; Martínez-Ortega et al. 2004). In the present study we analyze AFLP patterns for a broad sampling of *G. aridum* populations, including from Colima and several populations of all species of subsection *Erioxylum* and subsection *Integrifolia*. We examine levels and patterns of genetic diversity as revealed by AFLP analysis, address the taxonomic and phylogeographic implications of these data, and explore the data for evidence of historical, intersubsectional, nuclear introgression.

**Materials and Methods**

**Plant Material**

We studied 143 individuals from 50 populations of *Gossypium aridum* (Rose & Standley) Skovsted, *G. laxum* Phillips, *G. lobatum* Gentry, and *G. schwendimanii* (Fryxell & Koch) from subsection *Erioxylum* and of *G. davidsonii* Kellogg, *G. klotzschianum* Andersson from subsection *Integrifolia* (see Appendix, available online only at http://dx.doi.org/10.1554/05-184.1.s1). Plants were grown in the greenhouse at Iowa State University, starting with seed samples selected from collections made during previous collecting trips and those described elsewhere (DeJoode 1992; Wendel and Albert 1992; Wendel and Percival 1990), except for accessions 56-1 thru 66-5 (see online Appendix), which were provided by M. Ulloa and J. M. Stewart, and accessions 107-1, 118-1, 147-1, 157-1, 185-1, k3-1, and 32a-1, which were maintained as adult plants in the greenhouse. Sampling of *G. aridum* was designed to encompass its range of distribution (see Fig. 2), but in addition all populations from Colima were included. Following seed scarification, only 34% of seeds germinated and survived, and thus the initial 85 populations selected were reduced to 50 as follows: *G. aridum* (24, of which four are from Colima), *G. davidsonii* (10), *G. klotzschianum* (4), *G. laxum* (7), *G. lobatum* (2), *G. schwendimanii* (2), and *Gossypium* sp. (1) (see online Appendix for details).

One accession of a taxon suspected to represent an undescribed species (*Gossypium* sp.) also was included (acc. 64-1, US-72) (Álvarez et al. 2005; Ulloa et al. 2006). Vouchers for all individuals were deposited at the Ada Hayden Herbarium at Iowa State University, Ames.

**AFLP Analysis**

Fresh leaf tissue was used to isolate total DNA with the Plant DNAeasy kit (Qiagen, Valencia, CA) following the manufacturer’s instructions. Quality of isolated DNA was checked on 1% TAE-agarose gels. Genomic DNA (~200 ng) was digested with 10 units of EcoRI and 5 units of MseI, incubating at 37°C for 3 h. Double-stranded adaptors were prepared from the following complementary single-stranded oligonucleotides: 5’ CTC GTA GAC TGC GTA CC 3’ and 5’ AAT TGG TAC GCA GTC 3’ for the EcoRI adapter pair, and 5’ GAC GAT GAG TCC TGA G 3’ and 5’ TAC TCA GGA CTC AT 3’ for the MseI adapter pair. Ligation reactions were performed by adding 75 pmol of the EcoRI adapter, 15 pmol of the MseI adapter, and 20 units of T4 DNA ligase with its buffer to the digested product, and incubating overnight at 16°C. Ligation products were diluted by adding 160 μl of H2O.

For the preselective polymerase chain reaction (PCR), we added 10 μl of the diluted ligation product to 40 μl of a preselective PCR mix consisting of: 13 μl of H2O, 5 μl 10× PCR buffer, 1.5 μl MgCl2 (50 mM), 4 μl dNTP (2.5 mM), 8 μl (5 pmol/μl) of each preselective primer, and 0.5 μl of Fig. 2. Location of Mexican populations included in the study: *Gossypium aridum* (circles), *G. davidsonii* (squares), *G. laxum* (triangles), *G. lobatum* (crosses), *G. schwendimanii* (five-tip stars), *Gossypium* sp. (eight-tip stars). Colima populations are indicated by an arrow.
Taq DNA polymerase (5 U/μl). Sequences of preselective primers are: EcoRI + A: 5’ GAC TGC GTA CCA ATT CA 3’, and MseI + C: 5’ GAC GAT GAG TCC TGA GTA AC 3’. Preselective PCR conditions were a preliminary 75°C extension for 2 min followed by 20 cycles of 94°C for 30 sec, 56°C for 30 sec, 75°C for 2 min, finishing with 1 cycle of 60°C for 30 min. Ten μl of this PCR product were electrophoresed through 1.5% TAE-agarose gels and stained with ethidium bromide to verify adequate preselective amplification. The remaining 40 μl were diluted with 740 μl of dH2O.

For the selective PCR, we added 5 μl of diluted preselective PCR product to 20 μl of the selective PCR mix consisting of: 11.5 μl dH2O, 2.5 μl 10× PCR buffer, 0.75 μl MgCl2 (50 mM), 3 μl dNTP (2.5 mM), 0.75 μl (5 pmol/μl) each of two EcoRI labeled (6-FAM and TET) selective primers, 0.5 μl (50 pmol/μl) of one MseI unlabeled primer, and 0.25 μl of Taq DNA polymerase (5 U/μl). We performed selective PCR with eight primer pairs (see Table 1). The PCR profile was 1 cycle of 94°C for 2 min, 1 cycle of 94°C for 30 sec, 65°C for 30 sec, and 72°C for 2 min, followed by 9 cycles of a 1°C decrease in annealing temperature per cycle, followed by 35 cycles of 94°C for 30 min, 36°C for 30 sec, and 72°C for 2 min, and a final extension at 60°C for 30 min. For all samples we performed duplicates of preselective and selective reactions to verify reproducibility. Selective PCR products were electrophoretically separated using automated sequencing gels on an ABI Prism 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA) at the DNA Sequencing and Synthesis Facility at Iowa State University. See www.eeb.iastate.edu/faculty/WendelJ/aflp.htm for more details on our AFLP protocol.

**Scoring AFLP Data**

Gel images were analyzed with ABI GeneScan Analysis 3.0 and Genotyper 2.5 (Applied Biosystems) software. To verify reproducibility, samples were run in a manner that permitted duplicates to be visualized side by side on each gel. GeneScan row data was transformed into a Microsoft Excel spreadsheet with the aid of the program Genescan Step 3 (R. J. Dyer, pers. comm.). Peaks between 100 and 500 bp in length were scored, and only those with 100% reproducibility in all replicate samples were included in the data matrix. To avoid ambiguities, we eliminated peaks that overlapped more than half with others, peaks with relative fluorescence units below 100, and other possible gel artifacts. Our guiding philosophy was to use only the most robust, repeatable, and unambiguous bands. A presence/absence matrix was constructed from the final data table.

**Data Analysis**

Genetic diversity was estimated based on percentage of polymorphic bands within each taxon under consideration. To test for nuclear introgression within populations of *G. aridum* from Colima, we treated the latter as a different entity where appropriate. Thus, 10 entities were defined a priori for all analyses: *G. aridum* from Colima, *G. aridum* outside of Colima, *G. davidsonii*, *G. laxum*, *G. lobatum*, *G. klotzschianum*, *G. schwendimanii*, *Gossypium* sp., subsection *Erioxylum*, and subsection *Integifolia*. Genetic similarity (Dice coefficients) was calculated among and within entity (see Table 2). Polymorphic bands were classified into two main categories: exclusive (only present in one entity) and shared (present in more than one entity). We arbitrarily defined sub-categories within these two main categories. Thus, we classified exclusive bands as diagnostic (present in 100% of individuals of all populations of the entity), potentially diagnostic (present in at least 25% of the populations in the entity), and rare (present in less than 25% of the populations in the entity). Similarly, shared bands were categorized either as diagnostic (present in at least 25% of the populations of one entity and in less than the 25% of populations of any other entity) or not diagnostic. Based on this classification, genetic composition of each entity was estimated (see Table 3). In an attempt to visualize possible nuclear introgression into Colima populations, we represented the AFLP band distribution (see Fig. 3) for all shared bands present in Colima populations, including bands diagnostic for subsections (*Erioxylum* and *Integifolia*).

Analyses of molecular variance (AMOVA) at different tax-
TABLE 3. Distribution of diagnostic and potentially diagnostic fragments scored for each entity analyzed. Ni, number of individuals; Nf, number of fragments scored; Np, number of populations; p, percentage of polymorphic fragments within each entity; e, percentage of exclusive fragments; shC, percentage of fragments exclusively shared with Colima populations; shE, percentage of shared fragments diagnostic for subsection Erioxylum, shl, percentage of shared fragments diagnostic for subsection Integrifolia.

<table>
<thead>
<tr>
<th>Entity</th>
<th>Ni</th>
<th>Np</th>
<th>Nf</th>
<th>p</th>
<th>e</th>
<th>shE</th>
<th>shC</th>
<th>shl</th>
</tr>
</thead>
<tbody>
<tr>
<td>G. aridum exclusive of Colima</td>
<td>50</td>
<td>20</td>
<td>285</td>
<td>89.8</td>
<td>26.6</td>
<td>41.1</td>
<td>4.6</td>
<td>4.9</td>
</tr>
<tr>
<td>G. aridum from Colima</td>
<td>16</td>
<td>4</td>
<td>169</td>
<td>58</td>
<td>13.7</td>
<td>45.6</td>
<td>—</td>
<td>10.7</td>
</tr>
<tr>
<td>G. davidsonii</td>
<td>36</td>
<td>10</td>
<td>166</td>
<td>48.2</td>
<td>7.8</td>
<td>18.1</td>
<td>0.6</td>
<td>33.1</td>
</tr>
<tr>
<td>G. klotzschianum</td>
<td>9</td>
<td>4</td>
<td>134</td>
<td>27.6</td>
<td>0</td>
<td>9.7</td>
<td>0</td>
<td>41</td>
</tr>
<tr>
<td>G. laxum</td>
<td>22</td>
<td>7</td>
<td>201</td>
<td>80.1</td>
<td>20.9</td>
<td>46.5</td>
<td>1.5</td>
<td>3</td>
</tr>
<tr>
<td>G. lobatum</td>
<td>6</td>
<td>2</td>
<td>134</td>
<td>23.9</td>
<td>2.2</td>
<td>57.5</td>
<td>1.5</td>
<td>3</td>
</tr>
<tr>
<td>G. schwendimanii</td>
<td>3</td>
<td>2</td>
<td>159</td>
<td>50.3</td>
<td>9.5</td>
<td>51.6</td>
<td>1.2</td>
<td>4.5</td>
</tr>
<tr>
<td>Gossypium sp.</td>
<td>1</td>
<td>1</td>
<td>118</td>
<td>—</td>
<td>2.5</td>
<td>54.2</td>
<td>0.9</td>
<td>1.6</td>
</tr>
<tr>
<td>Subsection Integrifolia</td>
<td>45</td>
<td>14</td>
<td>166</td>
<td>—</td>
<td>39.7</td>
<td>10.8</td>
<td>3</td>
<td>—</td>
</tr>
<tr>
<td>Subsection Erioxylum</td>
<td>98</td>
<td>36</td>
<td>429</td>
<td>—</td>
<td>39.4</td>
<td>—</td>
<td>12.8</td>
<td>4.2</td>
</tr>
</tbody>
</table>

The percentage of polymorphic AFLP bands within an entity serves as an estimator of diversity, although in our case this is only heuristic as it may be biased by differences in sample size, and in the case of the new species (Gossypium sp.) it is not applicable since only one individual is included. The highest percentage of polymorphic fragments was observed in G. aridum from outside Colima (89.8%) and in G. laxum (80.1%), whereas the lowest averages were found in G. klotzschianum (58%) and G. davidsonii (48.2%).

From the eight selective primer pairs (see Table 1) and 143 individuals, 495 fragments ranging from 100 to 500 bp were scored, of which 28 (13.9%) were monomorphic. The number of fragments per individual ranged from 129 to 180, with an average of 151 fragments. The highest averages of number of fragments per individual were found in G. aridum from Colima and G. schwendimanii (both 162), and in G. laxum (158), whereas the lowest averages were found in G. klotzschianum (135) and G. davidsonii (145). Four identical AFLP patterns were found for the following three populations of G. davidsonii: 30 (individuals 30-2, 30-4); 44 (individuals 44-1, 44-5); and 46 (individuals 46-1, 46-2, 46-3, 46-4, 46-5, 46-6).

The percentage of polymorphic AFLP bands within an entity serves as an estimator of diversity, although in our case this is only heuristic as it may be biased by differences in sample size, and in the case of the new species (Gossypium sp.) it is not applicable since only one individual is included. The highest percentage of polymorphic fragments was observed in G. aridum from outside Colima (89.8% over 20 pops.) and G. laxum (80.1% over seven pops.). More moderate values were obtained for G. aridum from Colima (58% over four pops.), G. schwendimanii (50.3% over two pops.), and G. davidsonii (48.2% over ten pops.). The lowest levels of polymorphism were found in G. lobatum (23.9% over two pops.) and in the Galapagos Island endemic G. klotzschianum (27.6% over four pops.), which previously has been shown to harbor limited genetic diversity (Wendel and Percival 1990). For a more realistic approach, sampling was standardized applying a rarefaction technique (Coart et al. 2005) with the aid of the program AFLPDIV available at http://www.pierroton.inra.fr/genetics/labo/Software. This technique is only applicable for entities with a minimum sample size of three populations with at least three individuals per population; thus,
G. klotzschianum, G. lobatum, G. schwendimanii, and the new species were excluded from this estimation. In this case, the highest percentage of polymorphic loci was obtained for G. aridum from Colima (44.1%) and G. laxum (41.5%), while G. davidsonii and G. aridum from outside Colima present similar lower percentages (35.2% and 35.6%, respectively).

Averages of interspecific and intraspecific pairwise similarities using Dice coefficients are presented in Table 2. As expected, the highest values generally were obtained within entities, ranging from 0.718 to 1.000, and the lowest values were observed when entities from the two different subsections were compared, ranging from 0.474 to 0.515. The species most similar to each other are G. davidsonii and G. klotzschianum, which exhibit a similarity coefficient (0.923) that is comparable to those obtained among populations within species. Within subsection Erioxylum, the species pair with the highest similarity coefficient is G. laxum/Gossypium sp. (0.819). Notably, the pairwise comparison within species with the lowest similarity is G. aridum from Colima and G. aridum from elsewhere in the range of the species (0.622). These two entities, however, were equally similar to subsections G. aridum and G. laxum from Oaxaca (0.718). Notably, the pairwise comparison within species with the lowest similarity is G. aridum from Colima and G. aridum from elsewhere in the range of the species (0.622).

To explore for evidence of nuclear introgression, we tabulated the proportion of shared fragments with other species of Mexican arid polyspecies (data not shown). Exclusive fragments for other populations of G. aridum from Colima were treated as an entity distinct from the remainder of G. aridum (structure B in Table 4), the percentage of variation explained increases, with 65.4% of total variation attributable to interentity variation and 26% arising from among populations within entities. Comparable results are obtained when Colima populations are excluded from the analysis (structure D in Table 4); 65.4% and 27.4% of variation is explained by among- and within-species differences, respectively. In each of these analyses, there exists relatively less variation within populations (all <9%). To explore whether the increase in variation explained when Colima populations are excluded from G. aridum is unique to populations from that region, we performed AMOVA with structure B (see Table 4), sequentially removing or treating as separate entities populations of G. aridum from Guerrero, Jalisco, Michoacán, Oaxaca, Puebla, and Sinaloa (data not shown in Table 4). This analysis indicates that only populations from Colima have a notable effect on percentage of variation explained among groups (Fig. 4).

Analysis of variance also revealed that the AFLP data are geographically structured, an expected result given the allopatry among some of the genetically distinct species. Treating each region as a group (structure E in Table 4), 59.1% of the total genetic diversity is attributed to variation among regions, whereas 32.1% of variation arises from differences among populations within regions. Similar percentages were obtained when Colima populations were eliminated from the analysis (structure F in Table 4).

Several analyses that included only G. aridum were performed to explore population structure within this widespread species and the effect of Colima populations on its molecular variance distribution. When groups are defined by regions, including Colima populations (structure G in Table 4), the highest percentage of variation (51.6%) is accounted for by regions, followed by among populations within regions.
(35.3%) and within populations (13.1%). In contrast, when Colima populations are excluded (structure H in Table 4), the results are notably different, with the majority of variation arising from differences among populations within regions (52.5%) as opposed to among regions (34.4%). When Colima populations are treated as a separate entity from the remainder of G. aridum (structure I in Table 4), 53.5% and 36.7% of total molecular variance arises from among-entity and within-entity variation, respectively; thus, over half of the variance arises from genetic divergence of Colima populations from the rest of the species. Finally, when each population of G. aridum is treated as its own group, the proportion of total variance arising from interpopulation differentiation is similar when Colima populations are either included (85.4%, structure J in Table 4) or excluded (85.6%, structure K in Table 4) from the analysis, with equal variance within populations in each case (14.6% and 14.5%, respectively).

Ordination and Cluster Analyses

To illustrate relative similarities among entities in multidimensional genetic space, we performed principal coordinate analysis (see Fig. 5). The first three principal coordinates collectively explain 61.5% of the variance and permit clear delineation of eight clusters. The first axis, which accounts for 40.5% of the variance, clearly separates two groups corresponding to subsection Integrifolia (G. davidsonii and G. klotzschianum) and subsection Erioxylum (G. aridum, G. laxum, G. lobatum, G. schwendimanii, and Gossypium sp.). The second coordinate (11.9% of total variance) segregates clusters within subsection Erioxylum, with the largest group comprising populations of G. aridum from Guerrero, Jalisco, Michoacan, and Sinaloa, which collectively are very close to two populations of G. aridum from Puebla. Gossypium aridum populations from the southernmost part of the species range (from Oaxaca) are located in an intermediate position of the second axis. The most striking result, however, is the remarkable differentiation of populations of G. aridum from Colima, which ordinate at the opposite end of PCoA2 from the remainder of the species.

Additional similarities evident in the principal coordinate analysis include the high similarity between G. lobatum and G. schwendimanii (which are not wholly separated in this analysis), the evident separation of these two species from G. laxum, and the relative closeness of this latter species to the new and as yet unnamed species (Gossypium sp.). Also, the PCoA analysis fails to separate the closely related but geographically disjunct (Wendel and Percival 1990) species G. davidsonii and G. klotzschianum.

<table>
<thead>
<tr>
<th>Structure to test</th>
<th>Source of variation</th>
<th>df</th>
<th>PV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among populations</td>
<td>49</td>
<td>90.1</td>
<td></td>
</tr>
<tr>
<td>Within populations</td>
<td>93</td>
<td>9.9</td>
<td></td>
</tr>
<tr>
<td>Among populations within regions</td>
<td>43</td>
<td>35.3</td>
<td></td>
</tr>
<tr>
<td>Within populations</td>
<td>93</td>
<td>8.3</td>
<td></td>
</tr>
<tr>
<td>Among entities</td>
<td>7</td>
<td>65.4</td>
<td></td>
</tr>
<tr>
<td>Among populations within regions</td>
<td>42</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>Within populations</td>
<td>93</td>
<td>8.6</td>
<td></td>
</tr>
<tr>
<td>Among regions</td>
<td>8</td>
<td>59.1</td>
<td></td>
</tr>
<tr>
<td>Among populations within regions</td>
<td>41</td>
<td>32.1</td>
<td></td>
</tr>
<tr>
<td>Within populations</td>
<td>93</td>
<td>8.9</td>
<td></td>
</tr>
<tr>
<td>Among regions</td>
<td>7</td>
<td>57.6</td>
<td></td>
</tr>
<tr>
<td>Among populations within regions</td>
<td>38</td>
<td>34.5</td>
<td></td>
</tr>
<tr>
<td>Within populations</td>
<td>81</td>
<td>7.9</td>
<td></td>
</tr>
<tr>
<td>Among regions</td>
<td>6</td>
<td>51.6</td>
<td></td>
</tr>
<tr>
<td>Among populations within regions</td>
<td>18</td>
<td>35.3</td>
<td></td>
</tr>
<tr>
<td>Within populations</td>
<td>42</td>
<td>13.1</td>
<td></td>
</tr>
<tr>
<td>Among regions</td>
<td>5</td>
<td>34.4</td>
<td></td>
</tr>
<tr>
<td>Among populations within regions</td>
<td>14</td>
<td>52.5</td>
<td></td>
</tr>
<tr>
<td>Within populations</td>
<td>30</td>
<td>13.1</td>
<td></td>
</tr>
<tr>
<td>Among entities</td>
<td>1</td>
<td>53.5</td>
<td></td>
</tr>
<tr>
<td>Among populations within regions</td>
<td>22</td>
<td>36.7</td>
<td></td>
</tr>
<tr>
<td>Within populations</td>
<td>42</td>
<td>9.9</td>
<td></td>
</tr>
<tr>
<td>Among populations</td>
<td>23</td>
<td>85.4</td>
<td></td>
</tr>
<tr>
<td>Within populations</td>
<td>42</td>
<td>14.6</td>
<td></td>
</tr>
<tr>
<td>Among populations</td>
<td>19</td>
<td>85.6</td>
<td></td>
</tr>
<tr>
<td>Within populations</td>
<td>30</td>
<td>14.5</td>
<td></td>
</tr>
</tbody>
</table>

**Table 4**. Results of molecular variance analyses (AMOVA) for different entities and geographic regions. Different structures to test are indicated by a capital letter in bold (A–K), df, degrees of freedom; PV, percentage of variation. Each group of the structure to test is included within parentheses: ARI, G. aridum from other than Colima; DAV, G. davidsonii; KLO, G. klotzschianum; LAX, G. laxum; LOB, G. lobatum; NEW, Gossypium sp. (new species); SCH, G. schwendimanii; ARICol, G. aridum from Colima; ARIGue, G. aridum from Guerrero; ARI Jal, G. aridum from Jalisco; ARIMich, G. aridum from Michoacan; ARIOax, G. aridum from Oaxaca; ARIPue, G. aridum from Puebla; ARISin, G. aridum from Sinaloa; Baja, populations from Baja California; Col, populations from Colima; Gue, populations from Guerrero; Jal, populations from Jalisco; Mich, populations from Michoacan; Oax, populations from Oaxaca; Pue, populations from Puebla; Sin, populations from Sinaloa. In all cases P < 0.001.
As with the ordination analysis, cluster analysis (see Fig. 6) clearly reveals the main split between the two subsections Integrifolia and Erioxylum. Within subsection Integrifolia, populations of G. davidsonii and G. klotzschianum, from Baja California and Galapagos Islands, respectively, are intermingled, although some structure is evident in the figure. Within subsection Erioxylum, populations of G. aridum from Colima form a group clearly distinct from the remainder of the subsection, with high bootstrap support (100%). High support (100%) also exists for the G. laxum/new species grouping and for G. laxum by itself (87%), whereas the grouping of G. lobatum with G. schwendimanii attains modest bootstrap support (57%). Within G. aridum exclusive of Colima, populations from Oaxaca are clearly differentiated (100% bootstrap support). The G. aridum core group by itself has low bootstrap support, and although populations from different regions form clusters, there is not a clear overall geographic pattern. Within this core group, the main split separates populations from Sinaloa (88% bootstrap support), Puebla (100% bootstrap support) and some from Jalisco (no bootstrap support) from a second group that includes populations from Michoacan (100% bootstrap support), Guerrero (separated in two groups, each one with low bootstrap support), and Jalisco (100% bootstrap support).

**DISCUSSION**

**Genetic Diversity and Similarities of Gossypium aridum and Its Allies**

*Gossypium aridum* is the most morphologically variable American diploid cotton species, as well as the most geographically widespread, extending from northwestern to southwestern Mexico with marginal populations reported from as distant as the Mexican Atlantic coast (e.g., from Veracruz). Its aggregate range overlaps the distributions of the remaining species of subsection *Erioxylum*, which are endemic to central Guerrero and Michoacan (*G. lobatum* and *G. laxum*) or are more locally distributed (e.g., *G. schwendimanii* is restricted to northern and eastern *Infiernillo* in Michoacan and Guerrero, and the new species of *Gossypium* is known only from one locality near the Balsas river in Guerrero; Ulloa et al. 2006). Species of subsection *Erioxylum* have long been thought to be closely related inter se, and molecular phylogenetic analyses focused on this subsection (Small and Wendel 2000) support this contention, as well as a scenario of relatively recent and perhaps rapid radiation. In addition, interspecific hybrids among species of subsection *Erioxylum* are frequently observed in the field (Ulloa et al. 2006), consistent with their occasional sympatry. Amplified fragment-length polymorphism data reported here further document the high level of genetic diversity within *G. aridum*, as revealed by the proportion of polymorphic fragments (see Table 3) and ordination (see Fig. 5) and clustering (see Fig. 6) results. Most notable is the striking divergence of the Colima populations from the remainder of the species (see Table 2) and the otherwise high intraspecific and geographically structured diversity within *G. aridum*. This helps explain why *G. aridum* is so morphologically and genetically diverse as well as why phylogenetic relationships among these closely related and sympatric species have been difficult to discern.

The AFLP data and analyses reported here highlight several features of the structure of genetic diversity in subsection *Erioxylum*. First, there is a clear correspondence between taxonomy and geography, presumably due to the endemic distribution of many taxa as well as some level of geographic partitioning within *G. aridum*. This is evidenced most clearly in the principal coordinate depiction (Fig. 5), although it also is apparent in the NJ tree (Fig. 6). For example, populations...
from Oaxaca comprise a genetically distinct grouping, as do the Colima *G. aridum* populations. Also, the two endemic species from Michoacan (*G. lobatum* and *G. schwendimanii*) form a clade, as do the two endemics from Guerrero, *G. laxum* and the new species.

The AFLP data indicate that members of subsection *Erioxylum* (*G. aridum*, *G. laxum*, *G. lobatum*, and *G. schwendimanii*) contain higher levels of genetic diversity (in terms of percentage of polymorphic fragments) than do *G. davidsonii* and *G. klotzschianum* (subsection *Integrifolia*). We
might expect this, given the greater distribution of the former as well as morphological suggestions that this would be the case (see also Wendel and Percival 1990). However, in some more narrowly circumscribed regions (e.g., G. laxum and G. aridum from Colima) relatively high genetic diversity is observed. Historical patterns of hybridization and/or large long-term effective population sizes may account for high levels of genetic diversity within subsection Erioxylum species.

The geographic and taxonomic distribution of AFLP bands (presumed here to be mostly homologous, as opposed to the more unlikely scenario of multiple, spurious, homoplous characters) lends some insight into evolutionary history and the genesis of present patterns, particularly when considered in light of the principal coordinate and NJ analyses. Perhaps the most noteworthy result is the genetic uniqueness and relative isolation of G. aridum from Colima, which displays the highest percentage of exclusive AFLP fragments, is strikingly divergent with respect to genetic similarity (Table 2), and clearly is isolated from the remainder of G. aridum in both ordination (Fig. 5) and clustering (Fig. 6) analyses. AMOVA quantifies this isolation of the Colima G. aridum populations, in that including Colima populations leads to much higher interpopulational differentiation than when these populations are excluded.

The genetic distinctiveness of Colima populations of G. aridum has not been morphologically evident. Recent observations in the field reported by Ulloa et al. (2006) indicate that all coastal, foothill populations of G. aridum from Jalisco, Colima, Michoacan, and Guerrero are morphologically similar, and only populations from Oaxaca (showing the largest leaves in the species with dense and fine indumentum) and populations of slopes and canyons from Colima (presenting the smallest leaves and capsules in the species) are macro-morphologically different from the remaining G. aridum. With respect to the latter observation, we did not find any characters indicative of morphological differentiation when accessions were grown in the greenhouse. Accordingly, and despite the evidence for some degree of molecular genetic differentiation, we do not consider the Colima populations to be a distinct species.

The AFLP data appear to be quite powerful in characterizing patterns of genetic diversity in subsection Erioxylum. In addition to the issue of the isolation of Colima G. aridum populations, G. aridum from Oaxaca exhibits a relatively high proportion of exclusive AFLP fragments and forms a group distinct from the remaining G. aridum populations (see Figs. 3, 4). Similarly, populations of G. laxum, which also have a high percentage of exclusive AFLP fragments, appear in clustering and ordination analyses to comprise a relatively isolated evolutionary/genetic unit. In contrast, some taxa are relatively weakly differentiated from congers. For example, G. lobatum and G. schwendingianii are suggested to hybridize (Ulloa et al. 2006) based on observations in the field of intermediate morphologies and of sympatry; here, these two entities are revealed for the first time to be close relatives and perhaps sister species. Even more genetically intermingled are populations of G. klotzschianum and G. davidsonii. The former contains no exclusive AFLP fragments, and its populations interdigitate with those of its close relative G. davidsonii in both clustering and ordination analyses. This is also reflected in the estimate of genetic similarity, which was 0.923 for the species pair (Table 2). These data are consistent with earlier suggestions, based on allozyme data, that these two species are not genetically well-differentiated (Wendel and Percival 1990).

A New Species of Gossypium

Included in the dataset was a single population of a taxon in subsection Erioxylum found to be molecularly divergent (Feng et al. 2003) in a preliminary report and considered to be an undescribed species (Ulloa et al. 2006). This putatively new species was first collected by M. Ulloa and J. M. Stewart, who noted that although its morphology in its defoliated condition resembles G. aridum, mature leaves become palmately lobed, as in G. laxum. Phylogenetic analysis of nuclear gene sequences supported the concept that this entity was a new species (Álvarez et al. 2005). The AFLP data presented here support the distinctive nature of this new species and lend additional support to the possibility that it shares a most recent common ancestor with G. laxum, with which it is genetically most similar (see Table 2), closest in multidimensional space (see Fig. 5), and closest in the NJ analysis (see Fig. 6). In addition, both species inhabit the same region (Guerrero), as do some populations of G. aridum, suggesting an additional possibility that the new taxon is of hybrid origin. We caution that the foregoing is based on a single collection of what apparently is a rare taxon, for which, hence, additional collections become a priority.

Intersubsectional Introgression into Colima Populations of Gossypium aridum

Inspection of shared and diagnostic AFLP polymorphisms indicate that the genomic composition of all species within subsection Erioxylum includes a mix of species-specific markers and, more commonly, fragments that are shared among two or more of the species. Rarer are instances where members of section Erioxylum share AFLP fragments that otherwise are diagnostic of subsection Integrifolia (1.6–4.9%; Table 3). This low level of shared AFLP fragments may reflect trans-specific polymorphisms arising from lack of coalescence, homoplasy among bands that are genetically different but of similar sizes, or possibly interspecific gene flow. However, from a comparative standpoint, the level of shared polymorphism between G. aridum from Colima and subsection Integrifolia becomes more remarkable, in that fully 10.7% of the AFLP fragments from Colima G. aridum are diagnostic of the otherwise genetically distant subsection Integrifolia (Table 3). Moreover, 3% of the AFLP polymorphisms from Colima are shared exclusively with subsection Integrifolia (data not shown), a situation not found for other subsection Erioxylum members (see Fig. 3 for details). The observation that Colima populations share more than a two-fold higher percentage of fragments with subsection Integrifolia than does any other subsection Erioxylum taxon may be taken as evidence of intersectional hybridization. We view lack of coalescence as an unlikely alternative to introgression because we see no reason why this phenomenon should effect the genomic composition of Colima populations more than that of any other Erioxylum species.
Additional support for the hypothesis of introgression comes from earlier analyses of chloroplast genomes (DeJoode 1992; Wendel and Albert 1992), which demonstrated a shared cytoplasmic ancestry for Integrifolia and *G. aridum* populations from Colima (schematically illustrated in Fig. 1). However, previous phylogenetic analyses based on sequences of low-copy nuclear genes (Cronn et al. 1996; Small and Wendel 2000; Álvarez et al. 2005) have failed to confirm the existence of nuclear introgression, perhaps because sampling only three genes is unlikely to reveal low-level nuclear gene flow. In fact, if our estimation for the amount of introgression is true, only about 3% of the AFLP polymorphisms present in Colima populations are exclusively found in subsection Integrifolia, meaning a low probability (0.03) of finding introgressed genes randomly drawn genes. The distinctiveness of Colima populations as well as their position in ordination and clustering analyses (Figs. 5 and 6) might be interpreted as a result of introgression between Integrifolia lineage and some other unknown and presumably extinct species, but this is unlikely because the morphology of Colima populations lies clearly within the range of variation of *G. aridum* (Ulloa et al. 2006). In addition, gene sequence data for several independent nuclear markers support the inclusion of the Colima populations within *G. aridum* (Cronn et al. 1996; Small and Wendel 2000; Álvarez et al. 2005). A more plausible interpretation of the relative distinctiveness of the Colima populations is that they have experienced some level of local genetic local differentiation, much like other *G. aridum* from other regions, in addition to introgression from Integrifolia.

We note that fingerprinting techniques such as AFLP analysis sample both nuclear and chloroplast genomes, and hence the question naturally arises as to whether the introgression we infer here represents bona fide interspecific transfer of nuclear material or merely shared cpDNA fragments. Whereas definitive proof of the genomic origin of shared AFLP fragments would require isolation, cloning, and sequencing of the bands in question, a simple justification for the hypothesis of nuclear introgression emerges from algebraic considerations based on genome sizes (900 Mbp for the nuclear genome and 0.15 Mbp for the chloroplast genome). With quasi-random sampling of these genomes by AFLP techniques, we expect to observe one chloroplast fragment for every 6000 nuclear fragments on AFLP gels. Our odds, therefore, of having observed any cpDNA fragments from among the 495 bands scored is low, and becomes vanishingly small for multiple, shared cpDNA fragments. A similar algebraic argument applies to mitochondrial DNA. Thus, we conclude that most of the Colima fragments shared exclusively (3%) with Integrifolia members, and those diagnostic for Integrifolia but present in Colima populations (10.7%) reflect introgression of nuclear material (excluding fragments that cryptically are homoplasious and those that are peculiarly sympleiomorphic to this comparison).

According to previous authors using phenetic and cladistic approaches (Small and Wendel 2000; Álvarez et al. 2005), speciation of American diploid cottons took place following separation of the Gulf of California from what is now mainland Mexico, estimated to have occurred around 6 to 12 million years ago (Larson et al. 1968; Lonsdale 1991; Ferrari et al. 1999; Ingle 2001). If this scenario is correct, introgression occurred after these geological and cladistic events, and thus the cytoplasmic and nuclear gene introgression from the Integrifolia lineage (from Baja California) into what now are Colima cottons required dispersal of *G. davidsonii* from Baja California to the Colima coast in southwestern Mexico (an ocean voyage of $\sim$750 km), followed by interspecific hybridization using the immigrant as the seed parent and *G. aridum* as the pollen parent. This would presumably have been followed by repeated backcrossing of the hybrid, as female, into the paternal lineage, ultimately generating a nearly pure but introgressant *G. aridum* nuclear genome (and morphology) but in an alien cytoplasm. This scenario gains credibility from the predilection for transoceanic dispersal observed in wild *Gossypium* (reviewed in Wendel and Cronn 2003), and from the many other remarkable stories of odds-defying interspecific gene flow in the genus (Cronn and Wendel 2004).

From a population genetic perspective, the foregoing scenario is rendered more plausible by the likelihood that only one or a few individuals of *G. davidsonii* (or its close but extinct relative) crossed the Gulf of California and became established on the Colima coast, and thus pollination would have been mainly or exclusively by *G. aridum*. At maturity, capsules of *G. aridum* dehisce to passively release seed that is dispersed by gravity; thus, long-term retention of a locally introgressed cytoplasm may have been promoted by passive, local seed dispersal, even in the face of continued paternal gene flow from *G. aridum*. The fact that sufficient evidence of this nuclear introgression persists until this day suggests that the process was initiated relatively recently, perhaps during the Pleistocene following formation of the Gulf of California. In addition, the persistence of foreign genomic material may have been promoted by locally small population sizes, long generation times associated with the tree habit and/or, possibly, natural selection favoring introgressed genes or gene interactions. The fact that typical (nonintrogressant) populations of *G. aridum* have not been found in Colima provides correlative support for the possibility of selective maintenance of introgressed genes in these populations. We note that extensive de novo genomic variation has been demonstrated to be provoked by intergeneric hybridization in rice (Wang et al. 2005), as has extensive chromosomal rearrangement in hybrid sunflowers (Lai et al. 2005). Thus, although in many cases mutations and/or changes in chromosomal patterns may be disadvantageous for the hybrid, they might also play an important role for either the transfer or evolution of novel traits related to adaptation and speciation in natural populations.

Among the many unanswered questions raised by this story are those concerned with timing and the dynamics of the introgressive hybridization event, about which at this point we can only speculate. However, it is experimentally feasible to explore the genomic distribution of the introgressed material, a possibility made realistic by the existence of high-density genetic maps (Rong et al. 2004) for diploid and polyploid cotton. It also would be of interest to inventory the introgressant genes and genetic regions, through judicious cloning and sequencing approaches. Such an analysis serves as a necessary prelude to any attempt to appreciate the potential adaptive relevance of this interspecific introgression.
Finally, we note an important implication of the *G. aridum* history, namely, that cryptic and seemingly improbable interspecific introgression may be a more important process in angiosperm evolution than presently realized (Rieseberg and Wendel 1993; Rieseberg et al. 1995, 2003; Rieseberg 1997; Cronn et al. 2003; Cronn and Wendel 2004).

**Acknowledgments**

We thank D. DeJooede, E. Percival, J. M. Stewart and M. Ulloa for seed collections; R. Cronn, D. DeJooede, and J. M. Stewart for their comments on the manuscript; C. Aedo for map design; J. Fuertes and G. Nieto for help with the ARLEQUIN and NTSYS programs and for comments; R. Percifeld for technical assistance; and J. I. Yagüe for his help in the greenhouse. Research support was provided by the U.S. National Science Foundation and by the Spanish Ministry of Science, Education, and Sports.

**Literature Cited**


Small, R. L., and J. F. Wendel. 2000. Phylogeny, duplication, and intraspecific variation of Adh sequences in New World diploid...
of numerical classification. W. H. Freeman, San Francisco, CA.
Ulloa, M., J. M. Stewart, E. A. García-C., S. Godoy-A., A. Gaytan-
M., and S. Acosta-N. 2006. Cotton genetic resources in the west-
ern states of Mexico: in situ conservation status and germplasm
collection for ex situ preservation. Genet. Res. Crop Evol. doi:
10.1007/s10722-004-2988-0.
Reduction of species in the wild potato *Solanum* section *Petota*
series *Longipedicellata*: AFLP, RAPD and chloroplast SSR data.
Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M.
Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Za-
in rice induced by introgression from wild rice (*Zizania latifolia*
genus (*Gossypium*): character-state weighted parsimony analysis
of chloroplast-DNA restriction site data and its systematic and
Wendel, J. F., and R. C. Cronn. 2003. Polyploidy and the evolution-
Wendel, J. F., and A. E. Percival. 1990. Molecular divergence in
the Galapagos Island-Baja California species pair, *Gossypium*

Corresponding Editor: P. Soltis