Plant Syst Evol
DOI 10.1007/s00606-012-0615-7

SHORT COMMUNICATION

Assessing the monophyly of polyploid *Gossypium* species

C. E. Grover · K. K. Grupp · R. J. Wanzek · J. F. Wendel

Received: 31 August 2011 / Accepted: 15 February 2012

Abstract The origin and monophyly of the polyploid cotton (*Gossypium*) species has been largely accepted, despite the lack of explicit phylogenetic evidence. Recent studies in other polyploid systems have demonstrated that multiple origins for polyploid species are much more common than once thought, raising the possibility that *Gossypium* polyploids also had multiple origins, as postulated by some authors. To test the monophyly of polyploid cotton, we sequenced a 2.8-kb intergenic region from all diploid species belonging to the genome groups from which the polyploid originates. The resulting phylogenetic analyses strongly support a single origin of polyploid cotton involving a D-genome ancestor related to *Gossypium raimondii* and an A-genome ancestor that was sister to both extant A-genome species.

Keywords Cotton · *Gossypium* · Monophyly · Polyploidy · Phylogeny

Introduction

Polyploidy is a common and often recurrent phenomenon that has played a role in the evolutionary history of all angiosperms (Soltis et al. 2009; Jiao et al. 2011). Far more prevalent than once imagined, repeated rounds of polyploidy are inferred in the ancestry of many plants (Soltis and Soltis 1993, 1999; Jiao et al. 2011), reflecting a recurring cycle of genome doubling and subsequent fractionation. Thus, understanding the patterns and consequences of polyploidy becomes integral to understanding diversity and plant evolution. Complicating the study of polyploid evolution is the observation that many polyploid species appear to have multiple origins from the same or similar progenitors (Soltis et al. 2004; Soltis and Soltis 1993, 1999).

Gossypium is a monophyletic genus of approximately 45 diploid and five polyploid species distributed throughout the tropics and subtropics in arid and semi-arid regions of Africa–Asia, Australia, and the New World (Wendel and Cronn 2003). Diploids are divided into eight genome groups (designated A–G and K), from which a single polyploid genome group ("AD") originated from hybridization between A-genome and D-genome progenitors in the mid-Pleistocene (Wendel and Cronn 2003; Endrizzi et al. 1985). Substantial research into the origins and consequences of polyploidization in *Gossypium* has led to a large knowledge base regarding changes accompanying genomic merger and doubling (Adams et al. 2003, 2004; Adams and Wendel 2006, 2005; Chaudhary et al. 2009a, b; Cronn et al. 1996; Flagel et al. 2009; Grover et al. 2004, 2007, 2008; Hawkins et al. 2008, 2006; Hovav et al. 2008; Rapp et al. 2009; Small and Wendel 2002; Wendel et al. 1995), which are of particular interest in the case of the two polyploid species that were domesticated (*G. barbadense*, *G. hirsutum*) and which now provide most of the world's cotton fiber. The genetic, physiological, and phenotypic transformations that accompanied the independent domestication of these two species is of great interest, and inferences gained by comparing these two domestications rely on the correct identification of the origin of the polyploid species and their closest diploid model progenitors.

Electronic supplementary material The online version of this article (doi:10.1007/s00606-012-0615-7) contains supplementary material, which is available to authorized users.

C. E. Grover (✉) · K. K. Grupp · R. J. Wanzek · J. F. Wendel
Ecology, Evolution, and Organismal Biology,
Iowa State University, Ames, IA 50011, USA
e-mail: corrinne@iastate.edu

Published online: 11 March 2012

The origin of polyploid cotton was uncertain for a number of years, with different data sets and opinions suggesting either of the two extant A-genome species and up to four of the extant D-genome species as possible best models for the extinct polyploid parents (Wendel 1989; Wendel and Cronn 2003). Additionally, questions were raised about the monophyly of the polyploid species (Parks et al. 1975; Johnson 1975), noting that multiple origins could not be excluded. To date, the monophyletic origin of polyploid cotton has been widely assumed, and has been supported by limited evidence, including chloroplast restriction site data (Wendel 1989) as well as biogeographic considerations (i.e., the unlikely nature of multiple transoceanic dispersals). There has not, however, been an explicit test of monophyly using a reasonably robust data set based on nuclear sequences. While chloroplast restriction site data suggest that all five polyploid species arose from a single cytoplasmic source (Small and Wendel 1999; Wendel 1989), recent data from other plants demonstrating the prevalence of multiple polyploid origins (Soltis and Soltis 1993, 1999; Soltis et al. 2004; Tate et al. 2006) have reiterated the need for an explicit test of polyploid monophyly in *Gossypium*, which we present here.

Materials and methods

Sampling and sequencing

Representatives from each diploid species belonging to the two parental genome groups (with the sole exception of *G. schwendemanii*), as well as representatives from each allopolyploid species (Table 1), were obtained from our greenhouse collection, and DNA was extracted at the Iowa State University (ISU) DNA facility using the Autogen service. Primers (Table 2) were designed to amplify a 2.8-kb intergenic region between a putative leucine-rich repeat family protein (GenBank: AAT64029) and a putative pentatricopeptide repeat protein (GenBank: AAT64030), including the 3′ end of both genes, from the previously

Table 1 Taxa included in the present study

Genome	Species	Accession
A_1	*G. herbaceum* L.	Wagad
A_2	*G. arboreum* L.	AKA8401
AD_1	*G. hirsutum* L.	MAXXA
AD_1	*G. hirsutum* L.	TM1
AD_1	*G. hirsutum* L.	TX1046
AD_1	*G. hirsutum* L.	TX2094
AD_1	*G. hirsutum* L.	TX488
AD_1	*G. hirsutum* L.	TX51
AD_1	*G. hirsutum* L.	Wilkes Island, Wake Atoll
AD_2	*G. barbadense* L.	K101
AD_2	*G. barbadense* L.	GPS 52 OWZH-045
AD_2	*G. barbadense* L.	Pima S6
AD_3	*G. tomentosum* Nutt. ex. Seem	WT936
AD_4	*G. mustelinum* Miers ex Watt	Local lab accession
AD_5	*G. darwinii* Watt	PW-45
AD_6	*G. ekmanianum* Wittmack	Krapovickas Sejio #5
D_1	*G. thurberi* Todaro	D1-5
$D_{2\text{-}1}$	*G. armourianum* Kearny	D2-1-6
$D_{2\text{-}2}$	*G. harknessii* Brandegee	Local lab accession
$D_{3\text{-}d}$	*G. davidsonii* Kellog	Local lab accession
$D_{3\text{-}k}$	*G. klotzschianum* Andersson	D3 k-3
D_4	*G. aridum* (Rose & Standley) Skovsted	DRD185
D_5	*G. raimondii* Ulbrich	JFW accession
D_6	*G. gossypioides* (Ulbrich) Standley	D6-2
D_7	*G. lobatum* Gentry	DRD157
D_8	*G. trilobum* (DC.) Skovsted	Local lab accession
D_9	*G. laxum* Phillips	DRD107
D_{10}	*G. turneri* Fryxell	D10-3
Gk	*Gossypioides kirkii* (Mast.) J. B. Hutchinson	Local lab accession

Table 2 Primers used to amplify the intergenic region used for phylogenetic reconstruction

Primer name	Primer sequence
LeuRR-F4	GGCTCTTTGTGGTCACAGTGAGAAA
LeuRR-F4b	TGACTTGCGAATAATGACTAGGCT
LeuRR-F5	ATCCACCAACACCAACAACTTGGC
LeuRR-F6	AGTTCAATGATTCCTCCGCCTCCT
LeuRR-F7	CACAACGAGCAATCCCAAGCTTCA
LeuRR-FA	AAAGTAAGGTCACAGCAACCTGGC
LeuRR-R3	TCATCCATGACCAGGTTGCTGTGA
LeuRR-R3A	TCATCCATGGCCAGGTTGCTGTGA
LeuRR-R5	CCTGAGTTTCTTCACCTAGTCACC
LeuRR-R6	CAAAGTATCGTCTTCTGCGGCTAC
LeuRR-R7	GTTGTGGCATATACCTGTTGGACG
LeuRR-R8	GGAGTCTTCTGGTGATAGTGACTC
LeuRR-RA	GGCCATTTCATGACAATCACCGCA
LeuRR-RB	CCTTTAGCTAAGGCTGTTAGCAAGT
LeuRR-RC	AGGAGCATCATCACGGGCCAAATA
LeuRR-RD	TGAAGCTTGGGATTGCTCGTTGTG

sequenced cellulose synthase region (Grover et al. 2004) of *Gossypium* (GenBank: AY632359-60 and EU626442-44). For the diploid species, a total of three primer combinations were used under standard polymerase chain reaction (PCR) conditions to amplify the region of interest; for the polyploid species, additional primers were required to obtain the A-genome homoeologous regions due to a large insertion shared by the A and A-derived genomes (Table 2). PCR products were purified with the PureLink gel extraction kit (Invitrogen) or e-gel system (Invitrogen), and cloned with the FastLink ligation kit (Epicentre) using the pGEM T-easy vector (Promega) and Top10 chemically competent cells (Invitrogen). For the diploid species, a minimum of eight clones per primer set was sequenced with vector primers using the ABI BigDye terminator v3.1 chemistry and run on the ABI 3730XL located at the ISU DNA facility; for the polyploid species, a minimum of 16 clones per primer set was sequenced in the same fashion. Sequences were manually trimmed of vector sequence and aligned with MUSCLE (Edgar 2004), and consensus sequences were created in Bioedit (Hall 2007). Nexus files for the Bayesian and parsimony analyses were generated using ClustalW (Thompson et al. 2002).

Phylogenetic analysis

Models of nucleotide substitution and model parameters for phylogenetic reconstruction were selected by jModelTest version 0.1.1 (Guindon and Gascuel 2003; Posada 2008), and phylogenetic relationships were assessed using both maximum-parsimony [PAUP* 4.0 beta 10 (Wilgenbusch and Swofford 2003)] and Bayesian analyses [MrBayes version 3.1.2 (Altekar et al. 2004; Ronquist and Huelsenbeck 2003; Huelsenbeck and Ronquist 2001)]. The Bayesian analyses were conducted with the following parameters: three runs with four chains for 10 million generations and using a burn-in fraction of 25%. Two models of evolution were used; the first model, TIM3 + G, was selected by jModelTest and was implemented with the derived fixed transition frequencies (1.8843, 4.4734, 1.0000, 1.8843, 5.8712, 1.0000 for r_{AC}, r_{AG}, r_{AT}, r_{CG}, r_{CT}, r_{GT}, respectively) and fixed nucleotide frequencies (0.3448, 0.1847, 0.2029, 0.2676 for A, C, G, T, respectively). In addition, the basic general time reversible + gamma (GTR + G) model was selected for comparison. For maximum parsimony, a heuristic search with 1,000,000 random addition sequence replicates and with TBR branch swapping was used. These same two models were used for the parsimony reconstructions, again using parameters selected by jModelTest. In addition to the full alignment, a further two rounds of analyses, identical to the first, were performed using (1) only nongapped positions and (2) only nongapped positions and excluding regions of putative gene conversion (see "Results and discussion").

Results and discussion

A total length of 2,798 nucleotides (nt) of aligned sequences were analyzed for 45 *Gossypium* accessions (representing 21 species) and one outgroup species (Table 1). The region amplified was mostly intergenic; however, it included the 3′ end of both the previously predicted gene "putative leucine-rich repeat family protein" (GenBank: AAT64029) and the predicted gene "putative pentatricopeptide repeat protein" (GenBank: AAT64030), which were sequenced as a part of the larger cellulose synthase region of *Gossypium* (Grover et al. 2004, 2008). The absolute sequence length ranged from 2,434 nt in the D-genome homoeolog of *G. hirsutum* (Wilkes Island) to 2,764 nt in the A-genome species and A-genome homoeologs. This difference in size was due primarily to a 231-nt insertion that arose in the A-genome ancestor, and a 95-nt deletion in the D-genome homoeolog of *G. hirsutum* (Wilkes Island). Aside from those large indels, the size of the indels ranged from 1 to 12 nt in length, with average size of 2.6 nt and median of 1 nt. Within the ingroup, the sequence distance, excluding gaps, ranged from 0.000 to 0.020 (Supplementary Table 1).

Recent analyses have demonstrated that "gene conversion," or nonreciprocal homoeologous recombination, is possible between homoeologs in allopolyploid *Gossypium* (Salmon et al. 2010). Thus, we inspected our data set for evidence of this process, which has the potential to

confound phylogenetic reconstructions. Two putative sequence conversion events between homoeologs were detected. One sequence conversion event occurred in the A homoeolog of *G. tomentosum* (AD3), from aligned positions 67–137, and included two nucleotide conversions (G→A at position 67 and A→T at position 137). This conversion event occurred in the 3′ exon of the "putative pentatricopeptide repeat protein" (GenBank: AAT64030) in codon positions 1 and 2, respectively. Both conversion events led to an amino acid change, replacing glutamic acid with lysine at position 67 and asparagine with isoleucine at position 137. A second conversion event was detected in the D homoeolog of the *G. barbadense* (AD2) K101 accession, from aligned positions 1501–1649, which included a total of nine nucleotide conversions (Supplementary 1), all located in the intergenic space. In both cases, the PCR, cloning, and sequencing were repeated to verify the conversion. Because conversion events can influence the phylogeny by shifting the phylogenetic placement of the converted sequence to the base of the clade (McDade 1990, 1995), alignments were constructed and analyzed that either included or excluded the affected regions (Fig. 1).

Phylogenetic reconstructions were completed using both Bayesian (MrBayes) and maximum-parsimony (PAUP* 4.10b) methods on three alignment files: (1) the full alignment including gaps as missing data, (2) the alignment with gapped positions removed, and (3) the alignment with both gapped positions and regions of conversion removed. jModelTest 0.1.1 (Guindon and Gascuel 2003; Posada 2008) was used in all cases to determine the best-fit model of nucleotide substitution and to select the most appropriate model parameters (see "Materials and methods"); in addition, for all alignments and reconstruction methods, the general time reversible model + gamma distribution was also used as a generic model for comparison with the more specific model (i.e., TIM3 + G) selected by jModelTest.

Phylogenetic inferences of monophyly

Because allopolyploids contain two homoeologs, corresponding to the A-genome and D-genome donors, we generated two sets of sequences that resolve into two halves of each tree. With respect to the D-genome half, the American diploid cottons (subgenus *Houzingenia* Fryxell) are a monophyletic group of 13 species located primarily in southwest Mexico and extending north into Arizona. Phylogenetic relationships among the species have been evaluated (Alvarez et al. 2005), and previous research has identified *Gossypium raimondii* (D5) as the closest living relative of the paternal ancestor to all polyploid species using non-nuclear molecular methods and other evidence (Endrizzi et al. 1985; Wendel and Cronn 2003). In the present study, we used sequence data from an intergenic region to explicitly test the relationship of D-genome species to all allopolyploids. Both Bayesian and maximum parsimony clearly identified *G. raimondii* (D5) as the closest living relative of the ancestral D-genome donor for all of the polyploid species (Fig. 1), in agreement with previous evidence. For all reconstructions using both the GTR and TIM3 models, *G. raimondii* was confirmed as the closest living ancestor to the polyploid D-genome donor (Fig. 1). For both Bayesian and maximum parsimony, *G. raimondii* is either resolved as sister to the D-genome portion of the polyploid clade (Fig. 1a, c), or is the only D-genome diploid that is nested within that clade (Fig. 1b, d), the latter reflecting reduced resolution arising from eliminating positions containing gaps or converted nucleotides.

The ability of the present data to resolve relationships *among* the polyploid species and accessions is minimal, with most of the *G. hirsutum* (AD1) accessions forming a clade, and no clear resolution among other taxa (Fig. 1). Relationships among the remaining D-genome species achieve better resolution, in partial agreement with previous reports (Alvarez et al. 2005) and with certain strong associations. Subsection *Erioxylum* was recovered, as expected from prior sequencing, and *G. gossypioides* (D6) was retained as the basal-most D-genome diploid, as in previous data (Alvarez et al. 2005). The remaining subsections are in most cases, paraphyletic; however, section *Houzingenia* is monophyletic in the present study [*contra* (Alvarez et al. 2005)]. The differences between the present data and those previously reported are likely attributable to both rapid radiation and hybridization/introgression at the base of the D-genome cottons, which makes phylogenetic reconstruction with strong support less probable (Alvarez et al. 2005; Cronn et al. 2003). The aim of the present study, however, was not to resolve relationships among D-genome species, but instead to test the hypothesis of a monophyletic origin of the polyploid species.

With respect to the A-genome half of the trees, the two extant African A-genome cottons, *G. arboreum* and *G. herbaceum* (subgenus *Gossypium*), are equivalent models of the maternal genome donor to polyploid cotton (Wendel 1989), although phonetically some evidence suggests that *G. herbaceum* may more closely resemble the A-genome donor than *G. arboreum* (Wendel et al. 2010). All evidence suggests that the polyploid species were formed following a transoceanic dispersal of the A-genome ancestor to Mesoamerica, where the polyploid species formed and then subsequently spread to much of the Caribbean, Central America, northern South America, the Hawaiian Islands, and many other Pacific Islands (Fryxell 1979; Wendel and Cronn 2003; Wendel et al. 2010). Given the unlikely nature of multiple transoceanic dispersals of

more than one A-genome ancestor, it has been considered unlikely that there was more than one A-genome ancestor to the polyploid species. The present data lend molecular support to the inference that only one A-genome ancestor was involved in forming the polyploid species, and that the formation of the polyploid species occurred only once. In all analyses, both Bayesian and maximum parsimony, the two extant A-genome species cluster together and are sister to the polyploid species (Fig. 1). There is a notable exception, however, in the clustering of the diploid and polyploid A-genomes that is attributable to the observed large tract of sequence conversion in the A-genome of the polyploid *G. tomentosum* (AD3), as mentioned above. In both the Bayesian and maximum-parsimony generated trees (Fig. 1a, c), the A-genome sequence from AD3 is positioned at the base of the A-genome sequences, a result

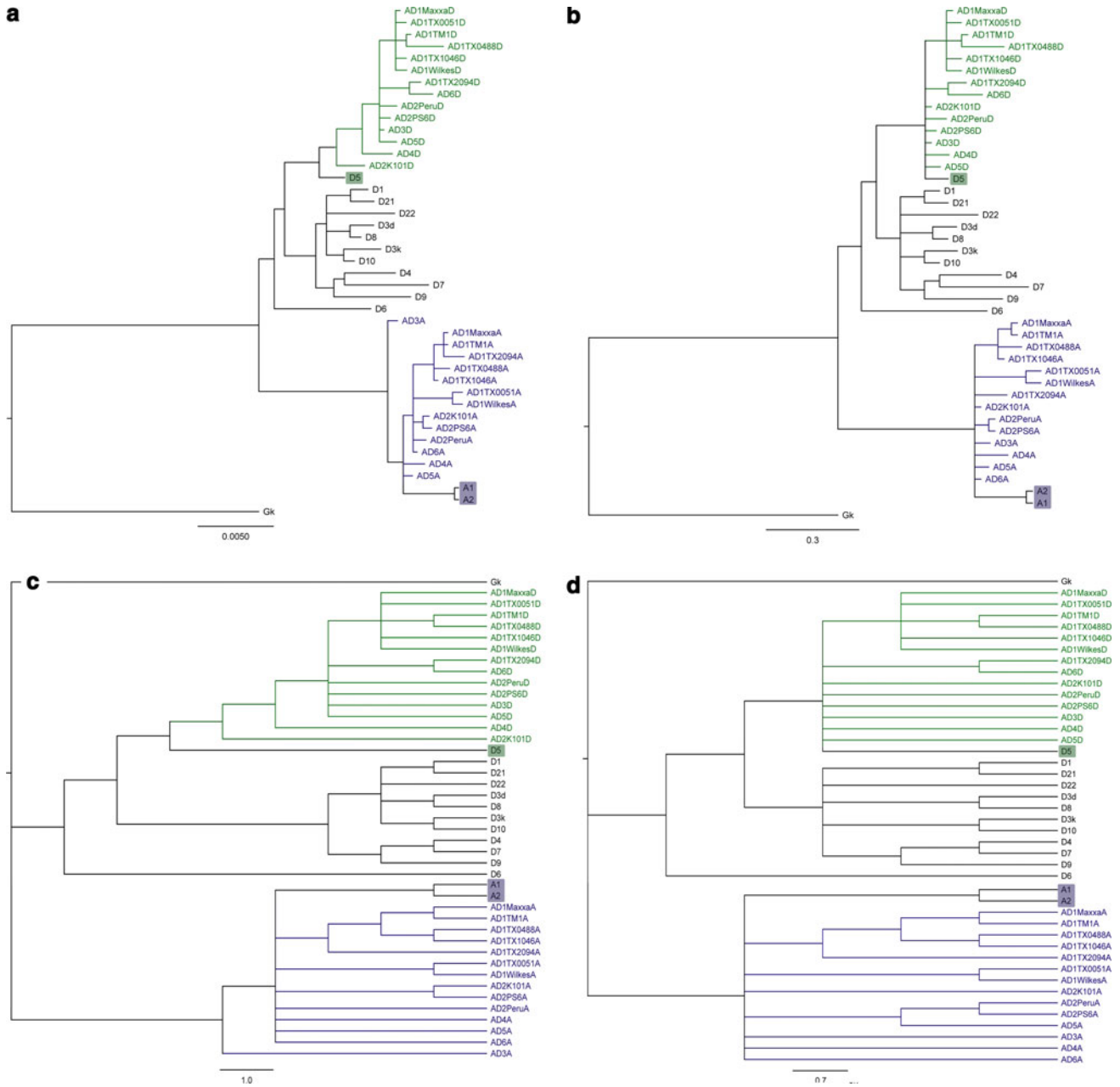

Fig. 1 Majority-rule consensus trees found by Bayesian and maximum-parsimony analyses. **a** Bayesian consensus tree for the full intergenic alignment, including gapped regions and regions involving sequence conversion. **b** Bayesian consensus tree created using an alignment from which gapped regions and regions involving sequence conversion in one or more species were removed. **c** Maximum-parsimony consensus tree for the full intergenic alignment, including gapped regions and regions involving sequence conversion. **d** Maximum-parsimony consensus tree created using an alignment from which gapped regions and regions involving sequence conversion in one or more species were removed

of the sequence conversion making AD3 seem more "D-like." Although the converted tract only resulted in two nucleotide changes, given the lack of sequence divergence between the A-genomes in general, this appears to be enough to pull the AD3 sequence toward the D-genome, consequently placing it at the base of the A-genome clade. This inference is supported by both the Bayesian and maximum-parsimony trees that were generated by the alignment that excluded those regions of sequence conversion (Fig. 1b, d), in which AD3 becomes nested within the largely unresolved A-genome clade, as expected.

Conclusions

The monophyletic origin of polyploid cotton has been largely accepted for years, despite the lack of an explicit molecular evaluation to support this supposition. Given recent evidence in many other taxa for repeated polyploid formation (Soltis and Soltis 1993, 1999; Soltis et al. 2004), it has become necessary to formally test that possibility for cotton. The data presented here demonstrate that the D-genome species *G. raimondii* and an A-genome species much like modern *G. arboreum* and *G. herbaceum* were involved in the creation of the polyploid species, and that it is this single combination that gave rise to the polyploids. Based on these data, and bolstered by the sheer biogeographic improbability of multiple transoceanic dispersals, we conclude that the polyploid *Gossypium* has a monophyletic origin.

References

Adams KL, Wendel JF (2005) Novel patterns of gene expression in polyploid plants. Trends Genet 21(10):539–543

Adams KL, Wendel J (2006) Allele-specific, bi-directional silencing of an alcohol dehydrogenase gene in different organs of interspecific cotton hybrids. Genetics 17:2139–2142

Adams KL, Cronn R, Percifield R, Wendel JF (2003) Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proc Natl Acad Sci USA 100(8):4649–4654

Adams KL, Percifield R, Wendel JF (2004) Organ-specific silencing of duplicated genes in a newly synthesized cotton allotetraploid. Genetics 168(4):2217–2226

Altekar G, Dwarkadas S, Huelsenbeck JP, Ronquist F (2004) Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. Bioinformatics 20(3):407–415

Alvarez I, Cronn R, Wendel JF (2005) Phylogeny of the New World diploid cottons (*Gossypium* L., Malvaceae) based on sequences of three low-copy nuclear genes. Plant Syst Evol 252(3–4):199–214

Chaudhary B, Flagel L, Stupar RM, Udall JA, Verma N, Springer NM, Wendel JF (2009a) Reciprocal silencing, transcriptional bias and functional divergence of homeologs in polyploid cotton (*Gossypium*). Genetics 182(2):503–517

Chaudhary B, Hovav R, Flagel L, Mittler R, Wendel JF (2009b) Parallel expression evolution of oxidative stress-related genes in fiber from wild and domesticated diploid and polyploid cotton (*Gossypium*). BMC Genomics 10:378

Cronn RC, Zhao X, Paterson AH, Wendel JF (1996) Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. J Mol Evol 42(6):685–705

Cronn R, Small RL, Haselkorn T, Wendel JF (2003) Cryptic repeated genomic recombination during speciation in *Gossypium gossypioides*. Evolution 57(11):2475–2489

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32(5):1792–1797. doi:10.1093/nar/gkh340

Endrizzi JE, Turcotte EL, Kohel RJ (1985) Genetics, cytology, and evolution of *Gossypium*. In: Caspari EW, John GS (eds) Advances in genetics, vol 23. Academic, London, pp 271–375

Flagel LE, Chen L, Chaudhary B, Wendel JF (2009) Coordinated and fine-scale control of homoeologous gene expression in allotetraploid cotton. J Hered 100(4):487–490 (pii: esp003)

Fryxell PA (1979) The natural history of the cotton tribe (Malvaceae, tribe Gossypieae), 1st edn. Texas A&M University Press, College Station

Grover CE, Kim H, Wing RA, Paterson AH, Wendel JF (2004) Incongruent patterns of local and global genome size evolution in cotton. Genome Res 14(8):1474–1482. doi:10.1101/gr.2673204

Grover CE, Kim H, Wing RA, Paterson AH, Wendel JF (2007) Microcolinearity and genome evolution in the *AdhA* region of diploid and polyploid cotton (*Gossypium*). Plant J 50(6):995–1006

Grover CE, Yu Y, Wing RA, Paterson AH, Wendel JF (2008) A phylogenetic analysis of indel dynamics in the cotton genus. Mol Biol Evol 25(7):1415–1428

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52(5):696–704

Hall T (2007) BioEdit: biological sequence alignment editor written for Windows 95/98/NT/2000/XP. http://www.mbio.ncsu.edu/bioedit/bioedit.html

Hawkins JS, Kim H, Nason JD, Wing RA, Wendel JF (2006) Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. Genome Res 16(10):1252–1261

Hawkins JS, Hu G, Rapp RA, Grafenberg JL, Wendel JF (2008) Phylogenetic determination of the pace of transposable element proliferation in plants: copia and LINE-like elements in *Gossypium*. Genome 51(1):11–18

Hovav R, Chaudhary B, Udall JA, Flagel L, Wendel JF (2008) Parallel domestication, convergent evolution and duplicated gene recruitment in allopolyploid cotton. Genetics 179(3):1725–1733

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17(8):754–755

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, Soltis DE, Clifton SW, Schlarbaum SE, Schuster SC, Ma H, Leebens-Mack J, dePamphilis CW (2011) Ancestral polyploidy in seed plants and angiosperms. Nature 473(7345):97–100

Johnson BL (1975) *Gossypium palmeri* and a Polyphyletic Origin of New World Cottons. Bull Torrey Bot Club 102(6):340–349

McDade L (1990) Hybrids and phylogenetic systematics I. patterns of character expression in hybrids and their implications for cladistic analysis. Evolution 44(6):1685–1700

McDade L (1995) Hybridization and phylogenetics. In: Hoch PC, Stephenson AG (eds) Experimental and molecular approaches to plant biosystematics. Missouri Botanical Garden, Missouri

Parks CR, Ezell WL, Williams DE, Dreyer DL (1975) The application of flavonoid distribution to taxonomic problems in the genus *Gossypium*. Bull Torrey Bot Club 102(6):350–361

Posada D (2008) jModelTest: phylogenetic model averaging. Mol Biol Evol 25(7):1253–1256

Rapp R, Udall J, Wendel J (2009) Genomic expression dominance in allopolyploids. BMC Biol 7(1):18

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19(12):1572–1574

Salmon A, Flagel L, Ying B, Udall JA, Wendel JF (2010) Homoeologous nonreciprocal recombination in polyploid cotton. New Phytol 186(1):123–134. doi:10.1111/j.1469-8137.2009.03093.x

Small RL, Wendel JF (1999) The mitochondrial genome of allotetraploid cotton (*Gossypium L.*). J Hered 90:251–253

Small RL, Wendel JF (2002) Differential evolutionary dynamics of duplicated paralogous Adh loci in allotetraploid cotton (*Gossypium*). Mol Biol Evol 19(5):597–607

Soltis D, Soltis P (1993) Molecular data facilitate a reevaluation of traditional tenets of polyploid evolution. Crit Rev Plant Sci 12:243

Soltis D, Soltis P (1999) Polyploidy: recurrent formation and genome evolution. Trends Ecol Evol 14:348

Soltis DE, Soltis PS, Tate JA (2004) Advances in the study of polyploidy since plant speciation. New Phytol 161(1):173–191. doi:10.1046/j.1469-8137.2003.00948.x

Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, dePamphilis CW, Wall PK, Soltis PS (2009) Polyploidy and angiosperm diversification. Am J Bot 96(1):336–348

Tate J, Ni Z, Scheen A, Koh J, Gilbert C (2006) Evolution and expression of homeologous loci in *Tragopogon miscellus* (Asteraceae), a recent and reciprocally formed allopolyploid. Genetics 173:1599

Thompson JD, Gibson TJ, Higgins DG (2002) Multiple sequence alignment using ClustalW and ClustalX. Curr Protoc Bioinformatics Chapter 2:Unit 2.3. doi:10.1002/0471250953.bi0203s00

Wendel JF (1989) New World tetraploid cottons contain Old World cytoplasm. Proc Natl Acad Sci 86(11):4132–4136

Wendel JF, Cronn RC (2003) Polyploidy and the evolutionary history of cotton. In: Advances in agronomy, vol 78. Academic Press, London, pp 139–186

Wendel JF, Schnabel A, Seelanan T (1995) Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). Proc Natl Acad Sci 92(1):280–284

Wendel JF, Brubaker CL, Seelanan T (2010) The origin and evolution of *Gossypium*. In: Stewart JM, Oosterhuis DM, Heitholt JJ (eds) Physiology of cotton. Springer, Dordrecht, pp 1–18

Wilgenbusch JC, Swofford D (2003) Inferring evolutionary trees with PAUP*. Curr Protoc Bioinformatics Chapter 6: Unit 6.4