A New Species of Cotton from Wake Atoll, *Gossypium stephensii* (Malvaceae)

Joseph P. Gallagher,¹ Corrinne E. Grover,¹ Kristen Rex,² Matthew Moran,³ and Jonathan F. Wendel¹,⁴

¹Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, Iowa 50011, U. S. A.
²Colorado State University, Center for Environmental Management of Military Lands in support of the US Air Force, 611 Civil Engineer Squadron, Fort Collins, Colorado 80523, U. S. A.
³USDA Forest Service, Chugach National Forest, Anchorage, Alaska 99501, U. S. A.
⁴Author for correspondence (jfw@iastate.edu)

Communicating Editor: Martin Wojciechowski

**Abstract**—Wake Atoll is an isolated chain of three islets located in the Western Pacific. Included in its endemic flora is a representative of the genus *Gossypium* colloquially referred to as Wake Island cotton. Stanley G. Stephens pointed out that “Wake Island cotton does not resemble closely either the Caribbean or other Pacific forms.” Taking into consideration morphological distinctions, the geographic isolation of Wake Atoll, and newly generated molecular data presented here, we conclude that the cottons of Wake Atoll do in fact represent a new species of *Gossypium*, here named *Gossypium stephensii*. This name is chosen to commemorate the eminent natural historian, evolutionary geneticist, and cotton biologist, S. G. Stephens.

**Keywords**—Chloroplast DNA, Pacific Ocean, phylogeny, targeted sequence capture, taxonomy.

Wake Atoll is an isolated chain of three islets (Peale, Wake, and Wilkes) located in the Western Pacific approximately 3,500 km west of Hawaii and 2,500 km east of Guam (Fig. 1; Levenson 2008). The remoteness of this island group in conjunction with its diminutive land area constrains its floristic and faunistic diversity, but also provides the opportunity for evolutionary endemism. The floristic communities of Wake Atoll have been cataloged by several naturalists, including E. H. Bryan Jr. and F. R. Fosberg (Bryan 1942; Fosberg 1959). Within the *Tournefortia* L. (Boraginaceae)-dominated forest found on the islets of the atoll, there is a representative of the genus *Gossypium* that has often been identified as *G. hirsutum* L., a member of the alloploid clade of cottons, and is colloquially referred to as Wake Island cotton (Fosberg 1959; Levenson 2008).

The Wake Atoll cotton specimens collected by Bryan and Fosberg were labeled *Gossypium hirsutum var. religiosum* (L.) G. Watt (Bryan 1942; Fosberg 1959). As Fryxell (1968) pointed out, the name *Gossypium religiosum* L. is “the most tortured of all the Linnean names in *Gossypium*” due to its complex history. While Linnaeus originally named *G. religiosum* in 1767, Watt (1907) reduced it to synonymy with *G. hirsutum*, after his examination of the original holotype. Fryxell (1968) notes that the holotype itself bears the annotation “in Indiis,” suggesting a specimen origin in either the West Indies or India. Thus, we conclude that this name has no bearing on the naming of the new entity described here.

Wake Atoll cotton was included in early biogeographic research on the genus. Stephens (1966) considered Wake Island cotton in his work on oceanic dispersal, treating it as a wild form of *G. hirsutum*. Notably, he pointed out that “Wake Island cotton does not resemble closely either the Caribbean or other Pacific forms,” noting its “sprawling shrub” growth habit, “densely hairy” pubescence, and larger than average petal spot in comparison to the other Pacific cottons (Stephens 1966). With consideration of prevailing ocean currents in conjunction with seed buoyancy and saltwater survival tests, Stephens suggested that Wake Island cotton may have originated following dispersal from the western coast of Mexico (Stephens 1966).

In our studies of these older, as well as more recent, collections made by members and support contractors of the United States Air Force, we validated the distinguishing characteristics of the Wake Atoll forms relative to *G. hirsutum* (Fryxell 1992). We further assessed its distinctiveness and taxonomic status through the use of extensive DNA sequence data from both the nuclear and chloroplast genomes, which reveal a clear differentiation of the Wake Atoll cottons from a large sampling of *G. hirsutum* accesions. Taking into consideration morphological distinctions, the geographic isolation of Wake Atoll, and newly generated molecular data presented here, we conclude that the cottons of Wake Atoll do in fact represent a new species of *Gossypium*, here named *Gossypium stephensii*. J. Gallagher, C. Grover, & Wendel. This commemorative name honors the late S. G. Stephens (1911–1986), an underappreciated and remarkably insightful natural historian, evolutionary geneticist, and cotton biologist (Wendel and Goodman 2011).

**Materials and Methods**

**Plant Materials**—Seeds collected on Wake Atoll by Ray Fosberg (formerly USGS, Washington, D.C.) were obtained from Paul A. Fryxell (formerly University of Texas, Austin) by J. Wendel. Additional collections were made by M. Moran and K. Rex. Plants were grown in the R.W. Pohl Conservatory at Iowa State University from these original seed collections. For comparative studies, we also sampled widely from the six other alloploid (AD genome) species of *Gossypium* (Grover et al. 2015), as well as from representative diploids that serve as models of the ancestral A- and D-genome donors (Appendix 1; see Wendel and Grover 2015). DNA was extracted from leaves of each accession using the Qiagen DNeasy Plant Mini Kit (Qiagen, Valencia, California), as per manufacturer instructions.

**Sequence Capture and Sequencing**—Sequencing libraries were constructed at West Virginia University Genomics Core Facility (http://genomics.as.wvu.edu/) using the Illumina TruSeq kit. Following library construction, DNA libraries underwent targeted sequence capture with custom bait sequence using the Mycroarray MYbaits kit (http://www.mycroarray.com/), as per the manufacturer’s protocol. Our custom bait sequence pool was designed to enrich 267 genes, as detailed in Appendix 2. Target-enriched libraries were then sent to either the Iowa State DNA Facility (http://www.dna.iastate.edu/) or the Beijing Genomics Institute (http://www.genomics.cn/en/index), where they were sequenced (150-bp paired end) on an Illumina HiSeq 2300.

**Sequence Assembly**—Reads were quality trimmed using sickle (Joshi and Fass 2011) and were mapped to the *G. raimondii* D-genome reference sequence (Paterson et al. 2012) for nuclear genes, or to the *G. hirsutum* chloroplast genome (Lee et al. 2006) for chloroplast genes, using GSNAP
Nuclear genes were mapped in combination with a diagnostic SNP-index (Page et al. 2013) and partitioned into A-genome and D-genome derived reads via PolyCat (Page et al. 2013). This process partitions reads into the two homoeologs expected for each gene in allopolyploid genomes. We treated *G. stephensii* as an allopolyploid due to its recognized morphological similarity, as well as previous molecular treatments of the plants (e.g. Grover et al. 2012). Both the chloroplast reads and the partitioned, mapped nuclear gene reads were assembled into individual sequences using bam2consensus from the BamBam suite of tools (Page et al. 2014). The resulting gene alignments were iteratively processed to remove uncaptured sequences/gene regions (process_alignments; https://github.com/Wendellab/phylogenetics), using a threshold of >80% ambiguous nucleotides (N) per sequence or >20% Ns per position for the first round and >50% Ns per sequence (30% for chloroplast) or >10% Ns per position for the second round. Any final gene or chloroplast alignment that did not contain at least one representative of each polyploid species, as well as a minimum of four Wake Atoll representatives, was excluded from the dataset. This quality filtering process resulted in the final inclusion of: (1) 102,227 bases of sequence from the chloroplast genome, roughly 65% of the total chloroplast genome [as compared to Lee et al. (2006)]; and (2) 474 genic regions representing 234 homoeologs from the A-genome and 240 homoeologs from the D-genome.

**Phylogenetic Analysis**—A maternal phylogeny was generated from the cpDNA sequences from a maximum likelihood analysis using RAxML (Stamatakis 2014, 2015). Parameters selected included rapid bootstrap analysis, GTR + Γ model of evolution, and 1,000 alternative starting trees. Trees were visualized using FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

For the nuclear data, alignments were filtered for PCoA analysis by first removing genes that did not contain all accessions sequenced, concatenated using SequenceMatrix (Vaidya et al. 2011), and subsequently removing all ambiguous sites. These concatenated sequences were then input into R for principal coordinates analysis (PCoA) using the ape package (Paradis et al. 2004; R Core Team 2015). Scripts
for implementing these analyses are available on https://github.com/Wendellab/phylogenetics.

**Results**

**Chloroplast Phylogenetics**—Chloroplast DNA sequences were assembled for representatives of all polyploid species of *Gossypium* (Appendix 1), including five representatives of *G. stephensii* and 34 representatives of *G. hirsutum*, the species to which *G. stephensii* was formerly thought to belong. A total of 102,227 aligned positions, representing ~65% of the *Gossypium* chloroplast genome, relative to *G. hirsutum* (Lee et al. 2006), were recovered and analyzed with maximum likelihood via RAxML and a GTR + Γ model (Stamatakis 2014, 2015). All accessions of *G. stephensii* were recovered as a monophyletic group that is distinct from both *G. hirsutum* and *G. ekmanianum* Wittm. (Fig. 2). The overall topology of the tree is consistent with previous reports (Grover et al. 2012, 2015), with the *G. stephensii* clade phylogenetically sister to the *G. ekmanianum* clade. Strong bootstrap values (>90%) support almost all the species clades. The one exception is the divergence between *G. ekmanianum* and *G. stephensii*; each clade has 16% and 44% bootstrap support, respectively (Fig. 2).

**Fig. 2.** Maximum likelihood tree based on analysis of maternally inherited chloroplast DNA sequences. The *Gossypium stephensii* clade is phylogenetically distinct from the other allopolyploid species, including *G. ekmanianum* and *G. hirsutum*. Node values are bootstrap support percentages.
Fig. 3. Principal coordinate analysis based on nuclear gene sequences from relevant allopolyploid and diploid *Gossypium* species. Separate analyses were conducted for homoeologs from the two co-resident genomes (A-, D-) in allopolyploid cottons. A. Axes 1 and 2 for the 237 concatenated A-genome gene sequences. B. Axes 1 and 2 for the 240 concatenated D-genome gene sequences. Key to species: AD1 = *G. hirsutum*; AD2 = *G. barbadense*; AD3 = *G. tomentosum*; AD4 = *G. mustelinum*; AD5 = *G. darwinii*; AD6 = *G. ekmanianum*; AD7 = *G. stephensii*. 
Analysis of Nuclear Gene Sequences—Because of the relatively recent diversification of the allopolyploid *Gossypium* clade and the attendant lack of species-level coalescence for many nuclear genes, phylogenetic analysis of different nuclear gene sequences yields multiple topologies (as described in Grover et al. 2015). Hence, to visually depict overall genomic relationships relevant to the taxonomic status of *G. stephensii*, we used a phenetic approach. Multiple accessions of *G. hirsutum* and *G. barbadense* L., as well as representatives from all other polyploid species (G. *tomentosum* Nutt. ex Seem., G. *mustelimum* Miers ex G. Watt, G. *darwinii* G. Watt, G. *ekmanianum*, G. *stephensii*; Appendix 1), were evaluated via PCoA of multilocus data (Fig. 3). Specifically, we included 474 genic regions, partitioned into 234 homoeologs from the A-genome (Fig. 3A) and 240 homoeologs from the D-genome (Fig. 3B), representing the two co-resident genomes of allopolyploid cotton (reviewed in Wendel and Grover 2015). Both analyses revealed similar overall depictions of genomic distinctions. Notably, the primary distinction for both sets of homoeologs along the first axis is the separation of the *G. hirsutum*-G. *ekmanianum*-*G. stephensii* grouping from the G. *barbadense*-G. *darwinii* species pair, with the representatives of G. *mustelimum* and G. *tomentosum* between the two. Along the second axis, which shows differentiation within the groups specified above, *G. stephensii* forms a tight cluster that is distinct from other species, and most importantly, it is clearly distinguished from a broad spectrum of *G. hirsutum* accessions spanning the full wild to domesticated continuum.

**Taxonomic Treatment**


*G. stephensii* differs from *G. ekmanianum* in having dense leaf pubescence, 8–11 bracteal teeth, subglobose capsules, and 3–4 locules in the capsule (Table 1). Although many characteristics of *G. stephensii* overlap with the species concept of *G. hirsutum*, it differs from *G. hirsutum* in lacking extrafloral nectaries at anthesis; these appear during capsule development (Table 1).

Sprawling woody shrubs, branching just above the base, with branches growing at nodes along woody stems; stems pubescent with stellate hairs; punctae black (lysigenous cavities; colloquially “gossypol glands”), abundant, more apparent in younger than in older stems. Leaves alternate, ascending, densely pubescent; stipules 6–10 mm; lamina cordate, weakly 3-lobed, rarely 5-lobed, broadly triangular to ovate, 65–75 mm long, 78–95 mm wide; apex acute to acuminate; margin entire, with stellate trichomes; adaxial and abaxial surfaces both with mostly stellate trichomes, pubescence denser on abaxial surface; 5 major veins abaxially raised; foliar nectaries on central midvein, 1.5 mm long, ± 3.5 mm from base; petals 50–53 mm long on mature leaves. Epicalyx 26–32 mm long, 26–28 mm wide, reflexed, dissected with 8–11 teeth, with numerous punctae throughout, diminishing toward the distal edge, hairs sparse, nectaries absent. Sepals 5–7 mm long, basally connate into a cup, lobes 1–2 mm long, acuminate, sinuses rounded, trichomes sparse to wanting, black punctae diminishing distally. Petals white to cream, with faint red to fuchsia basal spots, 39–42 mm long, petal spots 8–11 mm long. Staminodial column white; filaments 6–10 mm; pollen white to cream-colored. Style 18–20 mm long, exerted 11–12 mm beyond staminodial column; clavate, three ridged, punctate throughout; stigma decurrent on style. Capsules subglobose, apiculate, 18–20 mm in diam., with subtending epicalyx nectaries, locules 3–4, with 2–4 seeds per locule, capsule walls punctate. Seeds green, lanate, seed fibers tan, 10–12 mm long. Figure 4.

**Comments**—Because of the extensive and partially overlapping morphological variability between *G. stephensii*, *G. hirsutum*, and *G. ekmanianum* (Table 1; Fig. 5), DNA sequence data may be helpful for species confirmation. Within the chloroplast alignment, nucleotide changes specific to *G. stephensii*, as compared to the other allopolyploid species, appear at site 5,562 (A to T), within the intron of *psbD*; site 33,766 (T to A), in the intergenic space between *trnT-GGU* and *psbD*; and site 130,280 (A to C), in the intergenic space between *trnL-UAG* and *rpl32* (sites based on *G. hirsutum* chloroplast sequence; Lee et al. 2006).

**Distribution and Habitat**—*Gossypium stephensii* is found exclusively on the islets that make up Wake Atoll. Wake Atoll has a tropical maritime climate and is subject to trade winds and occasional typhoons. These plants occur in what has been described as a sparse *Tournesol argentea* L. f.-dominated forest. Cotton plants make up one of the more prevalent floristic elements, along with *Ipomoea pes-caprae* (L.) R. Br. These forests are found from the beaches up to the inland portion of the islands (Levenson 2008).

**Etymology**—The specific epithet honors the late Stanley George Stephens, paying tribute to his numerous contributions to our understanding of *Gossypium* diversity and evolution (Wendel and Goodman 2011). Stephens was a researcher at the Cotton Research Station in Trinidad and at North Carolina State University, where his many insightful studies of *Gossypium* evolution, systematics, and genetics

---

**Table 1.** Species diagnostic traits for *G. stephensii* and its most closely related species, *G. hirsutum* (based on Fryxell 1992) and *G. ekmanianum* Wittm. (based on Krapovickas and Seijo 2008).

<table>
<thead>
<tr>
<th>Feature</th>
<th><em>G. stephensii</em></th>
<th><em>G. hirsutum</em></th>
<th><em>G. ekmanianum</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth habit</td>
<td>Sprawling shrub</td>
<td>Dense</td>
<td>Scattered</td>
</tr>
<tr>
<td>Leaf pubescence</td>
<td>Dense</td>
<td>Dense to glabrous</td>
<td>Present</td>
</tr>
<tr>
<td>Floral nectaries</td>
<td>Absent</td>
<td>Present</td>
<td>Present</td>
</tr>
<tr>
<td>Bracteal Teeth</td>
<td>8–11</td>
<td>3–19</td>
<td>3–9</td>
</tr>
<tr>
<td>Petal size</td>
<td>40 mm</td>
<td>20–50 mm</td>
<td>40 mm</td>
</tr>
<tr>
<td>Capsule shape</td>
<td>Subglobose</td>
<td>Broadly ovoid</td>
<td>Broadly ovoid</td>
</tr>
<tr>
<td>Number of locules</td>
<td>3–4</td>
<td>3–5</td>
<td>3</td>
</tr>
<tr>
<td>Locality</td>
<td>Wake Atoll</td>
<td>Caribbean, Central America, South Pacific</td>
<td>Dominican Republic</td>
</tr>
</tbody>
</table>
Fig. 4. Holotype specimen of Gossypium stephensii J. Gallagher, C. Grover & Wendel.
Fig. 5. Illustrative morphological features of *Gossypium stephensii*: A. The sprawling growth habit of *G. stephensii* in its native habitat in Wake Atoll. B. Flower. C. Immature capsule. D. Dehisced capsule and seed, showing light reddish-brown fibers. E. Comparison of seed fibers from *G. hirsutum* cv. Maxxa, a domesticated line of Upland cotton (left), *G. stephensii* (center), and *G. hirsutum* TX2094, a wild form of *G. hirsutum* var. yucatanense (right).
resulted in his election into the National Academy of Sciences of the U. S. A. We name this species Gossypium stephensi in his memory.

Paratypes—U. S. A., Iowa: R. W. Pohl Conservatory, Bessey Hall, Iowa State University, grown from seed collected on Wilkes Island, received by J. F. Wendel from P. A. Fryxell in 1988. 5 May 2015, Gallagher 1 (ISC); R. W. Pohl Conservatory, Bessey Hall, Iowa State University, grown from seed collected on Peale Island, Wake Atoll, N19°18′33.2″E166°37′43.6″, collected by Matt Moran in August 2012. 31 October 2015, Gallagher 4 (ISC); R. W. Pohl Conservatory, Bessey Hall, Iowa State University, grown from seed collected on Peale Island, Wake Atoll, by Kristen Rex in March or April 2013. 31 October 2015, Gallagher 5 (ISC); R. W. Pohl Conservatory, Bessey Hall, Iowa State University, grown from seed collected on Peale Island, Wake Atoll, by Kristen Rex in March or April 2013. 31 October 2015, Gallagher 6 (ISC); R. W. Pohl Conservatory, Bessey Hall, Iowa State University, grown from seed collected on Peale Island, Wake Atoll, by Kristen Rex in March or April 2013. 31 October 2015, Gallagher 7 (ISC).

Acknowledgments. We thank Kara Grupp, Anna Tuchin, and Lori Schmidt for assistance in collecting plant materials for this project, and Paul A Fryxell for providing seeds. This work is funded by the National Science Foundation Plant Genome Program and by Cotton Incorporated. Joseph P. Gallagher is supported by NSF Graduate Research Fellowship DGE1247194.

Literature Cited


Appendix 1. alphabetical list of species considered in chloroplast phylogeny and principal coordinates analysis: species name, alphabetical list of accesses examined, asterisk if included in chloroplast phylogeny, voucher number, SRA accession number. All herbarium specimens are maintained at, ISC, with the exceptions of GB0262, GB027, GB0369, GB0319, LKT511, and TAMCOT, which perished prior to vouchering.

Gossypium arboreum L., A2-101*, Gallagher 14, SRR5001776; Gossypium barbadense L., GB0262*, no voucher, SRR5001789; GB028*, no voucher, SRR5001786; GB0303*, Gallagher 19, SRR5001784; GB0369*, no voucher, SRR5001799; GPPS2*, Gallagher 18, SRR5001801; K101*, Gallagher 17, SRR5001832; Phy76*, Gallagher 16, SRR5001793; Pima 56*, Gallagher 15, SRR5001811; Gossypium darwinii G. Watt, lab accession*, Gallagher 22, SRR5001813; Gossypium eckmanianum Wittm., TX2263*, SRR5001816; Gossypium barbadense L., TX2265*, SRR5001825; TX2266*, Gallagher 25, SRR5001814; TX2271*, Gallagher 26, SRR5001794; TX2273*, Gallagher 27, SRR5001775; Gossypium herbaceum L., A1-Wagad*, TX234, SRR5001795; Gossypium hirsutum L., ARK240*, Gallagher 12, SRR5001770; Cascot 1*, Gallagher 9, SRR5001823; Coker 315*, Gallagher 13, SRR5001827; CB2 252*, Krush and Grupp sp. (SIC450783), SRR5001826; FM 958* and Krush and Grupp sp. (SIC450786), SRR5001773; GB0319*, no voucher, SRR5001817; LKT511*, no voucher, SRR5001818; Maxxa*, Gallagher 8, SRR5001774; PM145*, Krush and Grupp sp. (SIC448082), SRR5001783; TAMCOT*, no voucher, SRR5001785; TM1*, Gallagher 10, SRR5001831; TX44*, Schmidt sp. (SIC448044), SRR50018102; TX48*, Schmidt sp. (SIC446777), SRR5001788; TX665*, Schmidt sp. (SIC446834), SRR5001822; TX672*, Schmidt sp. (SIC446832), SRR5001792; TX786*, Schmidt sp. (SIC446829), SRR5001803; TX1009, Schmidt sp. (SIC446828), SRR5001790; TX1037*, Schmidt sp. (SIC446782), SRR5001815; TX1046*, Schmidt sp. (SIC446820), SRR50018093; TX1055*, Krush and Grupp sp. (SIC448110), SRR5001780; TX1077*, Krush and Grupp sp. (SIC448120), SRR5001791; TX1110*, Krush and Grupp sp. (SIC450784), SRR5001798; TX1120* and Krush and Grupp sp. (SIC448122), SRR5001812; TX1162*, Krush and Grupp sp. (SIC448067), SRR5001809; Krush and Grupp sp. (SIC448091), SRR5001771; TX1122*, Krush and Grupp sp. (SIC448166), SRR5001796; TX1265*, Krush and Grupp sp. (SIC448043), SRR5001810; TX1748*, Krush and Grupp sp. (SIC47904), SRR5001795; TX1982*, Krush and Grupp sp. (SIC47907), SRR5001829; TX1988*, Krush and Grupp sp. (SIC448052), SRR5001797; TX1996*, Krush and Grupp sp. (SIC448073), SRR5001807; TX2002*, Krush and Grupp sp. (SIC450774), SRR5001830; TX2089*, Krush and Grupp sp. (SIC447866), SRR5001800; TX2109*, Krush and Grupp sp. (SIC448053), SRR5001772; TX2092*, Krush and Grupp sp. (SIC448055), SRR5001828; TX2094*, Krush and Grupp sp. (SIC448044), SRR5001804; TX209*, Krush and Grupp sp. (SIC447885), SRR5001787;