

Evolution of the Cotton Genus, *Gossypium*, and its Domestication in the Americas: A Review

Christopher R. Viot¹ and Jonathan F. Wendel²

¹UMR AGAP Institut - Univ Montpellier, CIRAD, INRAE, Institut Agro - F-34398 Montpellier, France. christopher.viot@cirad.fr

²Department of Ecology, Evolution, and Organismal Biology (EEOB), Bessey Hall, Iowa State University, Ames, IA 50011, USA. jfw@iastate.edu

Citation: Christopher R. Viot & Jonathan F. Wendel, 2023. Evolution of the Cotton Genus, *Gossypium*, and Its Domestication in the Americas. *Critical Reviews in Plant Sciences*, <https://doi.org/10.1080/07352689.2022.2156061>. Published online: 09 Jan 2023.

ABSTRACT	2
KEY WORDS	2
I - INTRODUCTION	2
II - EVOLUTION OF <i>GOSSYPIMUM</i> AND ORIGINS OF THE AMERICAN CLADES	3
A. GEOGRAPHIC DISTRIBUTION, TAXONOMY AND GENOME GROUPS	3
B - EMERGENCE OF THE AMERICAN CLADES	6
C - PHYLOGENY OF AMERICAN DIPLOIDS, THE SUBGENUS <i>HOUSINGENIA</i>	9
D - ALLOPOLYPLOIDIZATION AND EVOLUTION OF THE ALLOPOLYPLOID CLADE, SUBGENUS <i>KARPAS</i>	11
1. Age of the <i>Gossypium</i> allopolyploid lineage	12
2. Location of the allopolyploidization event	13
3. Diversification and geographical dispersal of the allopolyploid clade	14
III - DOMESTICATION OF THE TWO CULTIVATED ALLOPOLYPLOID COTTON SPECIES	15
A - ARCHAEOLOGICAL DATA ON TEXTILE USE OF COTTON IN MESOAMERICA AND SOUTH AMERICA	16
1. Mesoamerica and <i>G. hirsutum</i> cotton archaeology	16
2. South America and <i>G. barbadense</i> cotton archaeology	17
B – TAXONOMY OF THE TWO DOMESTICATED ALLOPOLYPLOIDS	18
C - DOMESTICATION OF <i>GOSSYPIMUM BARBADENSE</i>	18
1. Timeline and location of <i>G. barbadense</i> domestication	18
2. Intraspecific taxa in <i>G. barbadense</i>	19
a. <i>G. barbadense</i> var. <i>brasiliense</i>	21
b. <i>G. barbadense</i> type Sea Island	21
c. Egyptian and Pima ELS Cottons	22
D - DOMESTICATION OF <i>GOSSYPIMUM HIRSUTUM</i>	22
1. Genetic diversity in <i>G. hirsutum</i> and the concept of geographic or morphological races	22
2. Timeline and location of <i>G. hirsutum</i> domestication	25
E – POST-DOMESTICATION INTERSPECIFIC INTROGRESSION BETWEEN <i>GOSSYPIMUM HIRSUTUM</i> AND <i>G. BARBADENSE</i>	28
IV - CONCLUSIONS	29
ACKNOWLEDGMENTS	30
DECLARATIONS	30
COPYRIGHTS	30
REFERENCES	30
SUPPLEMENTARY DATA	39

Abstract

Gossypium, the cotton genus, includes ~50 species distributed in tropical and subtropical regions of all continents except Europe. Here we provide a synopsis of the evolutionary history of *Gossypium* and domestication of the American allopolyploid species, integrating data from fundamental taxonomic investigations, biogeography, molecular genetics, phylogenetic analysis and archaeology. These diverse sources of information provide a temporal and phylogenetic perspective on diversification among the diploids and on polyploid formation, uncover multiple previously cryptic interspecific hybridizations, clarify and contribute to the taxonomy of the genus, and offer a firm foundation for understanding parallel domestications in Mesoamerica and South America, which led to the globally important cotton crop species *G. barbadense* and *G. hirsutum*. *Gossypium* thus offers a testimonial example of the importance and utility of fundamental botanical discovery combined with modern technological capabilities to generate genomic insights into evolutionary history. We also review the current state of our knowledge regarding the archaeological history of cotton domestication and diffusion in the Americas, a seemingly unlikely story entailing parallel domestication origins and parallel directional selection tracing to 8,000 (*G. barbadense*) and 5,500 (*G. hirsutum*) years ago, transforming two geographically isolated wild short-day perennial shrubs having small capsules and seeds covered by short, tan-colored epidermal trichomes into modern daylength-neutral annuals bearing abundant, fine, strong white fibers. This dual domestication was followed several millennia later by unintentional and more recently intentional interspecific introgression, as the two species came into contact following their initial domestication in different hemispheres. Thus, the cycle of species divergence and biological reunion was reiterated, this time at the allopolyploid level. Understanding this evolutionary history is vitally important to our understanding of the genomic architecture of the world's most important fiber plant, and contributes substantially to our understanding of general biological principles.

Key words: phylogenetics; polyploidy, hybridization, domestication; archaeology; *Gossypium hirsutum*; *Gossypium barbadense*

I - Introduction

The cotton fiber used in textile production is harvested from plants belonging to the genus *Gossypium* L. Cotton is the most important natural textile plant in the world, providing 81% of global natural fiber and 27% of the world's textile production (FAO, 2021). Two allopolyploid species, *Gossypium hirsutum* L. and *G. barbadense* L., were independently domesticated in the American subtropics (Yuan *et al.*, 2021) and now are grown on all continents, producing the majority of the global cotton harvest. In addition, its oleoproteaginous seeds make cotton a major global crop for oil production and livestock feed (Bellaloui *et al.*, 2015). Because of the importance of cotton in clothing, in world industrial and economic development (Beckert, 2014; Wendel and Grover, 2015), and in the history of civilizations, the genus *Gossypium* has attracted much scientific interest from taxonomists, geneticists, archaeologists, ethnologists and historians.

In the past century, prior to the advent and application of modern molecular genetics, taxonomic and phylogenetic studies based on morphology, geography, artificial interspecific hybridization and cytogenetics led to the recognition of most modern *Gossypium* species and to a reasonably stable taxonomy (Beasley, 1940; Fryxell, 1979, 1992; Percival *et al.*, 1999; Wendel and Cronn, 2003; Konan *et al.*, 2009). This collective effort also generated an understanding that the genus naturally is divided into eight diploid "genome groups" (A to G and K) and one allopolyploid genome group designated AD which combines the African-Asian genome A with the American genome D (Wendel and Albert, 1992; Seelanan *et al.*, 1999; Wendel and Grover, 2015; Hu *et al.*, 2021). In addition to this broad, genus-level understanding, a great deal has been learned about diversity in both domesticated species, particularly in recent years, including insights into the nature of the wild forms, infraspecific classification, and interspecific gene flow (Hutchinson *et al.*,

1947; Hutchinson, 1951; Percy and Wendel, 1990; Wendel *et al.*, 1992; Brubaker *et al.*, 1993, 1999; Westengen *et al.*, 2005; Yuan *et al.*, 2021).

Archaeological and anthropological studies are the key to uncovering the timeline, location and human context of cotton domestication, the developments of its textile use and the geographic dispersal of its cultivation in the Antiquity. Millennia-old remains of plant-based textiles are generally seldom found as they can be preserved only in some precise conditions (Reynaud *et al.*, 2020), perhaps in South America in the dry environments of the Pacific coasts and in cold, high altitude sites in the Andes, and in Mesoamerica in a few sites of mountainous regions (Follensbee, 2008). Recent studies have pushed back the dates when cotton fibers first appeared in human settlements in the Americas to the eighth millennium BP¹ and a much clearer knowledge has been acquired about the history of cotton domestication (Splitstoser *et al.*, 2016). While other vegetal fibers were being used thousands of years earlier than cotton as textile material by humans in the Americas, at least from the 11th millennium BP (Jolie *et al.*, 2011), cotton appears to have been linked with the development of intensive activities in agriculture, fishing, craftsmanship, trade and, more globally, with the emergence of civilizations (Stark *et al.*, 1998; Pearsall, 2008; Beresford-Jones *et al.*, 2018), and it appears to have expanded like no other textile fiber, becoming nearly ubiquitous in tropical America by several thousand years ago.

In recent decades, molecular genetics and genomics studies have vastly expanded our understanding of the evolutionary history of *Gossypium* and the domestication process, illuminating with exquisite precision many details of interspecies relationships, timing of divergence events, the specific status of several natural population systems, and the surprisingly high frequency of mysterious interspecific gene flow and transfer of cytoplasmic organelles in the ancestry of modern species that would seem to have no opportunity to experience genetic contact. Here we synthesize data from disparate disciplines ranging from archaeology to genome sequencing to present a synopsis of, in part I, the evolution and phylogeny of the genus *Gossypium*, with particular emphasis on the American clades and, in part II, our understanding of domestication and patterns of diversity in the two American domesticated allopolyploids, *G. barbadense* and *G. hirsutum*. Our goal is to provide an accessible resource for those interested in the natural and human history of these vitally important plant species.

II - Evolution of *Gossypium* and origins of the American clades

A. Geographic distribution, taxonomy and genome groups

The genus *Gossypium* L. belongs to the subfamily Malvoideae within the large plant family Malvaceae, which has about 4225 species (Wang *et al.*, 2021). Within the Malvoideae, *Gossypium* and eight other genera constitute the tribe Gossypieae (Fryxell, 1979; Seelanan *et al.*, 1999; Phupathanaphong, 2006), also known as the cotton tribe, commemorating its most famous genus. The tribe Gossypieae is morphologically characterized by the distinctive synapomorphy of so-called “gossypol glands”, actually lysigenous pigment glands (Fryxell, 1968) containing concentrated terpenoids and providing defense against pests and pathogens (Gao *et al.*, 2020). All molecular phylogenetic studies support the monophyly of *Gossypium* (Seelanan *et al.*, 1997; Grover *et al.*, 2015a; Wendel and Grover, 2015; Hu *et al.*, 2021), and have shown, for example, that some taxa of uncertain generic placement are indeed in *Gossypium*. For example, analysis of a portion of chloroplast gene *ndhF* confirmed that *G. trifurcatum* (section Serrata), a species from Somalia, belongs to *Gossypium* (Rapp *et al.*, 2005), in spite of its serrated leaves, which is a trait not otherwise found in *Gossypium* and is, in fact, more characteristic of the related genus *Cienfuegosia*. The 52 presently accepted species of the genus *Gossypium* (Supplementary material S1) are organized into four subgenera, eight sections and ten subsections (Table 1), following the most recent and thorough taxonomic classification (Fryxell, 1979, 1992), a synthesis based on generations of workers using the many tools of classical taxonomy, including morphology, cytogenetics and geographical distribution.

Notwithstanding this wealth of understanding derived from decades of study, taxonomic uncertainty remains for certain parts of the genus, especially for populations from the Somali Peninsula (Horn of Africa) and the Arabian Peninsula (all presumably having E genomes). Also remarkable is that even though

¹ Before Present in calendar years, referring to 1 January 1950 as the commencement date of the time scale. Elsewhere in this paper, 'YBP' will stand for 'years BP' and 'MYA' (million years ago) for 'million years BP'.

Gossypium has been intensively studied, new species continue to be described, even among the allopolyploids. For example, the new species *G. ekmanianum*, from the Dominican Republic on Hispaniola, was initially described in 1928 but had since repeatedly been reclassified into *G. hirsutum* (Grover *et al.*, 2015b); chloroplast and nuclear gene sequences show unique polymorphisms and numerous diagnostic indels that distinguish *G. ekmanianum* from the other polyploid species and also indicate that it is a sister species to *G. hirsutum* (Grover *et al.*, 2015b). Similarly, *G. stephensii*, from the tiny coral atoll of Wake Island, far out in the Pacific Ocean, was long considered to be a form of *G. hirsutum*, but analyses of chloroplast and nuclear gene sequences demonstrated that this should be recognized as a new species distinct from both *G. hirsutum* and *G. ekmanianum* (Gallagher *et al.*, 2017). *Gossypium stephensii* appears closer to *G. ekmanianum* with respect to chloroplast DNA data, but to *G. hirsutum* using nuclear DNA data. Counterbalancing these additions to allopolyploid species diversity is a case of taxonomic lumping; using both chloroplast and nuclear DNA markers, *G. lanceolatum* was shown to not merit species status, but instead, that it merely is a mutant-leaved form of *G. hirsutum*, namely *G. hirsutum* race *palmeri* (Brubaker and Wendel, 1994).

Classical cytogenetics, i.e. karyotyping and studies of genomic affinity, have characterized eight genome groups (Konan *et al.*, 2009) for the diploid species and one for the allopolyploid species (Table 1). The cytogenetic classification resulted largely congruent with the classification and evolutionary history based on morphological traits and geographical distribution (Wendel *et al.*, 2010; Hu *et al.*, 2021).

With respect to cytogenetics and the concept of genome groups, perhaps the single most noteworthy scientific insight into *Gossypium* evolution was set in motion by the demonstration that the American cultivated species are tetraploids, with chromosome numbers of $2N=4X=52$ (Denham, 1924; Zaitsev, 1928)². More precisely, these species are allopolyploids, combining genomes from two different diploid ($2N=2X=26$) genome groups, the A and D (Skovsted, 1934).

Table 1. Subgenera, sections and subsections of *Gossypium*, with numbers of species in parentheses, and corresponding genome groups and genome sizes. Sources: Fryxell, 1979; Hendrix, 2005; Wendel *et al.*, 2009; Grover *et al.*, 2015b; Gallagher *et al.*, 2017. The precise genome assignment of each species is provided in Supplementary Material S1.

Genus	Subgenus	Section	Subsection	Genome group	Genome size
<i>Gossypium</i> (52)	<i>Gossypium</i> (14)	<i>Gossypium</i> (13)	<i>Gossypium</i> (2)	A	1667-1746 Mb
			<i>Anomala</i> (3)	B	1345-1359 Mb
			<i>Pseudopambak</i> (4-7)*	E	1496-1663 Mb
			<i>Longiloba</i> (1)	F	1311 Mb
		<i>Serrata</i> (1)	-	E?/B?	
	<i>Houzingenia</i> (13)	<i>Houzingenia</i> (7)	<i>Houzingenia</i> (2)	D	841-851 Mb
			<i>Caducibracteolata</i> (3)		856-910 Mb
			<i>Integrifolia</i> (2)		880-910 Mb
		<i>Erioxylum</i> (6)	<i>Erioxylum</i> (4)		919-934 Mb
			<i>Austroamericana</i> (1)		880 Mb
			<i>Selera</i> (1)		841 Mb
	<i>Sturtia</i> (18)	<i>Sturtia</i> (3)	-	C	1951-2015 Mb
		<i>Hibiscoidea</i> (3)	-	G	1756-1834 Mb
		<i>Grandicalyx</i> (12)	-	K	2450-2778 Mb
	<i>Karpas</i> (7)	-	-	AD	2347-2489 Mb

*Three species known only from herbarium specimens (Wendel *et al.*, 2009) are suspected to be E-genome taxa, but this has not been confirmed using modern tools.

All wild *Gossypium* species are distributed in arid to semiarid regions of the tropics and subtropics (Wendel and Grover, 2015) on all continents except Antarctica and Europe (Figure 1). *Gossypium* species are extraordinarily diverse in morphology, from herbaceous perennials to small shrubs to tree-like plants, with the majority being shrubs (Fryxell, 1979; Wendel and Grover, 2015). Four *Gossypium* species with longer hairs on their seeds have been domesticated and are cultivated for the production of the textile fiber

² Zaitsev (1928) cited in Fryxell (1979) p. 166

known as cotton: the diploid A-genome species *G. herbaceum* L. and *G. arboreum* L., and allopolyploid AD-genome species *G. hirsutum* L. and *G. barbadense* L.; these species are discussed in detail below. Using a combination of phylogenetic analysis and molecular clock-based dating (Wendel and Albert, 1992; Wendel and Grover, 2015; Hu *et al.*, 2021), we now understand that early during the evolution of *Gossypium*, approximately 7 to 5 MYA, the genus split into three main lineages (Figure 1), i.e., Australian, African, and American. Although it has not been conclusively demonstrated, Africa seems to be the most likely place where the genus *Gossypium* emerged, based on the highest diversity of species and genome groups from that continent as well as the observation that one of the two genera that comprise the phylogenetic outgroup to *Gossypium* (Wendel and Albert, 1992; Seelanan *et al.*, 1997), *Gossypioides*, also is African. From this African origin, long-distance dispersals are implicated in the spread to Australia and to the American tropics and subtropics. All of these long-distance dispersals occurred millions of years ago, that is, prior to the evolution of humans. Only much later, during the last 3,000 to 8,000 years, were the two African-Asian diploid and two American allopolyploid cotton species domesticated.

The worldwide dispersion of *Gossypium* species is noteworthy, as it must have required multiple trans-oceanic dispersal events. The most plausible hypothesis for this is some form of oceanic drift of either the seeds or perhaps mature capsules (Wendel and Grover, 2015), noting that seeds of many of these species appear tolerant of prolonged periods of immersion in salt water (Stephens, 1958, 1966). Long-distance dispersal of cotton seeds could also have entailed birds blown off course during major storms; in this respect the use of the hairy seeds for nest building has been observed (Molinet, 1692; Fryxell, 1992)³, and one of us (jfw) has seed of the Galapagos Island endemic *G. darwinii* collected from a native finch's nest. It also is possible, but perhaps less so, that storms themselves disperse seeds, without an avian or oceanic vector: storms that form over the Saharan desert can reach the Caribbean Sea in four days, carrying millions of tons of dust (Francis *et al.* 2020), potentially lifting into the air seeds of long-fiber African cottons, as hairy seeds plausibly could facilitate wind dispersal (Alavez *et al.*, 2021).

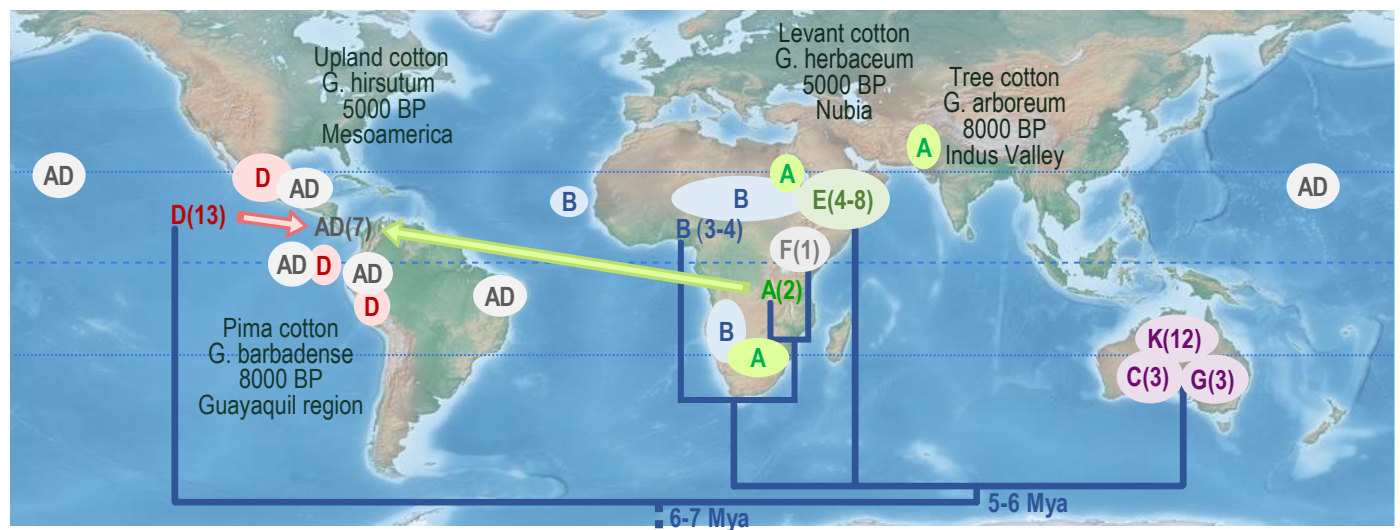


Figure 1. Genomic diversification and geographical dispersion of *Gossypium*, hypothetical regions of domestication, and times of earliest use of fibers from the four cultivated cotton species. Sources: Cronn *et al.*, 2002; Wendel *et al.*, 2009; Gallagher *et al.*, 2017; Viot, 2019. Letters designate genomic groups, with numbers of species in parentheses. Time in years before present.

A remarkable feature of the genus *Gossypium* is that genome sizes are strikingly variable (Table 1), ranging from the relatively small genomes of the American D-genome species to the nearly threefold-larger genomes of species from the Kimberley region of NW Australia. This is especially notable because all diploids share the same complement of chromosomes ($2n=26$). Modern genome sequencing studies all demonstrate that gene numbers are mostly conserved across genome groups and that the chromosomes are largely syntenic (Chen *et al.*, 2020), lending support to the now clear understanding that the vast majority of

³ Molinet (1692) cited in Jeanson (2020), p. 120

genome size variation in the genus arises from the differential proliferation and clearance of different families of different classes of transposable elements (Huang *et al.*, 2020), particularly gypsy-like retrotransposons (Hawkins *et al.*, 2009). *Gossypium* serves as an excellent example of the realization that genomes don't have a "one-way ticket to genomic obesity" (Bennetzen and Kellogg, 1997), but in fact, that reductions of TE content and hence genome size may be common in plant evolution in general (Leitch and Leitch, 2013), and in the *Gossypieae* as well (Wendel *et al.*, 2002). Intriguingly, the two genera comprising the phylogenetic outgroup to *Gossypium*, i.e., *Kokia*, from Hawaii, and *Gossypioides*, from East Africa and Madagascar, both have very small genomes characterized by extensive loss of TEs as well as genic content (Grover *et al.*, 2017).

B - Emergence of the American clades

Figure 2 summarizes our phylogenetic understanding of the diploid *Gossypium* genome groups, as well as the origin and diversification of the AD-genome allopolyploid clade, in the context of several representatives of both close (*Kokia*) and sequentially more distant outgroups (*Bombax*, *Theobroma*) within the Malvaceae, *sensu lato*, along a logarithmic time scale. Inside the two *Gossypium* clades native to tropical America, some extant species are indigenous to Pacific archipelagos and are denoted in Fig. 2 by a distinctive colour.

Malvaceae fossils, molecular clocks and appearance of American diploids and allopolyploids

Plant fossil studies may provide dates for the earliest appearance of plant orders, families and even some subfamilies; in the present application, data exist for the order Malvales, family Malvaceae and subfamily Malvoideae (Carvalho *et al.*, 2011; Richardson *et al.*, 2015; Hernández-Gutiérrez and Magallón, 2019). Combined with DNA sequence data, ages of plant fossils permit a calibration of the rates of genic synonymous (K_s) and non-synonymous (K_a) mutations (r = nucleotide substitutions/DNA site/year). Fossil finds related to Malvaceae include fossil leaves of Malvoideae from the middle-late Paleocene (58-60 MYA) Cerrejón Formation in Colombia (Carvalho *et al.*, 2011), indicating a divergence of the Byttnerioideae subfamily (*Theobroma cacao*) more than 60 MYA. Pollen attributed to the subfamily Bombacoideae – sister to subfamily Malvoideae – is dated from 66 to 72 MYA (Richardson *et al.*, 2015). Fossil leaves of Malvaceae and segregate families from the Late Cretaceous (100–66 MYA) give a computed 92 MYA for the stem node of Malvales (Magallón and Castillo, 2009). It should be noted that the only fossils for *Gossypium* are leaf prints dated to ca. 0.4 MYA from Hawaiian volcanic sediments (Cronn *et al.*, 2002), presumably from the Hawaiian endemic *G. tomentosum*.

Recent publications (Chen *et al.*, 2016; Pan *et al.*, 2020) have inferred molecular clock estimates between c. 60 MYA and 76 MYA for the *Gossypium*-*Theobroma* divergence, in agreement with fossil records. The estimated rates of synonymous (K_s) mutations (substitutions · synonymous site⁻¹ · yr⁻¹) for nuclear loci have varied somewhat, but are similar, e.g., $r = 3.48 \times 10^{-9}$ (Chen *et al.*, 2020), 3.61×10^{-9} (Grover *et al.*, 2017) or, in an earlier publication, 2.6×10^{-9} (Senchina *et al.*, 2003). The rate for non-synonymous substitutions (K_a) is about an order of magnitude lower than that of synonymous substitutions (Grover *et al.*, 2019), as expected for these more constrained sites. The rate for synonymous substitutions in nuclear loci has been approximated as 2.5-fold more rapid than for chloroplast sequences from the same species (Cronn *et al.*, 2002), although this estimate needs updating using modern genome sequences.

The evolutionary divergence between *Gossypium* and its closest outgroups, *Kokia* and *Gossypioides*, was first dated with molecular sequence data to ca. 12.5 MYA (Seelanan *et al.*, 1997), or to between 11.3 and 14.2 MYA (Senchina *et al.*, 2003), and to between 10.3 and 11.2 MYA (Grover *et al.*, 2017); this last estimate was obtained using a *Gossypium*-*Theobroma* divergence at 60 MYA for the molecular clock calibration, while whenever 70 MYA, a time closer to what fossil studies suggest (Richardson *et al.*, 2015; Hernández-Gutiérrez and Magallón, 2019), had instead been used for the calibration, the age for the divergence of *Gossypium* from *Kokia* and *Gossypioides* would have been computed as between 12.0 and 13.1 MYA. Thus, these estimates are quite consistent, providing a reasonably reliable inference of the temporal dimension of diversification in *Gossypium*.

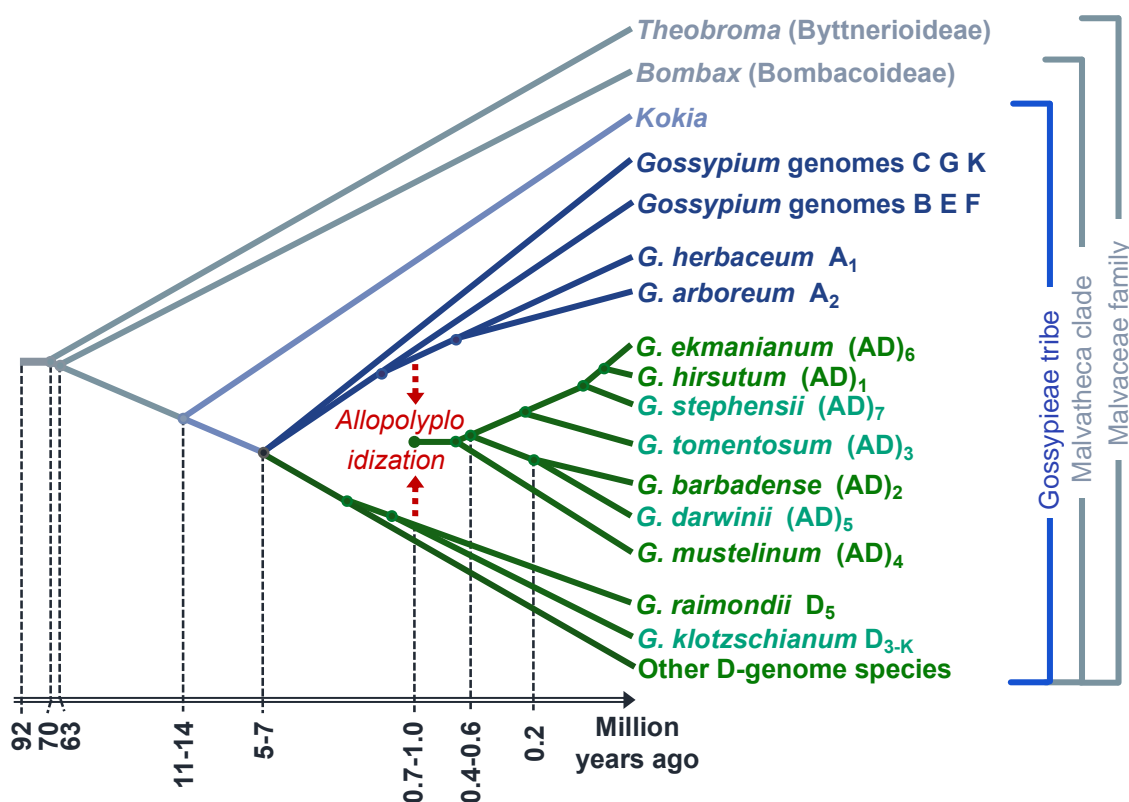


Figure 2. Schematic phylogenetic tree of *Gossypium*. Dotted arrows indicate the diploid nuclear genomes that were combined during allopolyploidization. Species in green are indigenous to the Americas; those in blue-green are indigenous to various Pacific Ocean archipelagos. Subfamily names are given in parentheses for genera outside of the subfamily Malvoideae. Logarithmic time scale in millions of years. Sources: Magallón and Castillo, 2009; Wendel *et al.*, 2009; Grover *et al.*, 2015a, 2015b; Magallón *et al.*, 2015; Gallagher *et al.*, 2017; Hernández-Gutiérrez and Magallón, 2019; Pan *et al.*, 2020.

The time of the basal partition of *Gossypium* into its three diploid subgenera representing the African, American and Australian clades (Figure 1) was initially estimated to be between 5 and 10 MYA (Wendel and Albert, 1992). One hypothesis stated that there was an early rapid radiation after *Gossypium* separated from the *Kokia-Gossypioideae* clade, with the major lineages having diverged by approximately 11 MYA (Wendel and Cronn, 2003; Wendel *et al.*, 2009; Grover *et al.*, 2017). Most molecular clock studies now estimate this basal divergence to have occurred ca. 5 to 7 MYA (Li *et al.*, 2015; Hu *et al.*, 2019; Chen *et al.*, 2020; Huang *et al.*, 2020) (see phylogenetic tree in Figure 1). This is about half the time since the divergence of *Gossypium* from the *Kokia-Gossypioideae* clade at approximately 11 to 12 MYA. We note that the genera *Gossypioideae* and *Kokia*, now in southeast Africa and the Hawaiian Islands, respectively, diverged from each other about 5.3 MYA (Grover *et al.*, 2017), close to when *Gossypium* split into its major lineages and dispersed over diverse continents.

Extensive phylogenetic effort has been expended elucidating relationships among the *Gossypium* diploid genome groups. As summarized in Figure 3, these attempts have not yielded an entirely congruent set of inferences regarding the branching orders near the base of the genus. Four main hypotheses have been forwarded (Fig. 3, panels A-D), differing from each other in several respects. All agree that the A and F genomes are close sister clades and that the three Australian genome groups (C, G and K) collectively constitute a monophyletic clade (Wendel and Grover, 2015; Chen *et al.*, 2016). The phylogenetic trees in Figure 3 differ, though, on whether the D-genome clade or the C+G+K-genomes clade diverged first and whether the B-genome species and/or the E-genome species are closer to the African A+F-genomes clade or to the Australian C+G+K-genomes clade. These alternative branching orders group themselves according to the type of DNA that is studied; plastid DNA (chloroplast DNA, or cpDNA) analyses tend towards the C+G+K-genomes clade diverging first and grouping with the B-genome clade (phylogenetic trees B and C in Figure 3), while nuclear DNA (nuDNA) analyses favor a hypothesis where the D-genome clade diverged

first and the E-genome clade is closer to the C+G+K-genomes clade (phylogenetic trees A and D in Figure 3).

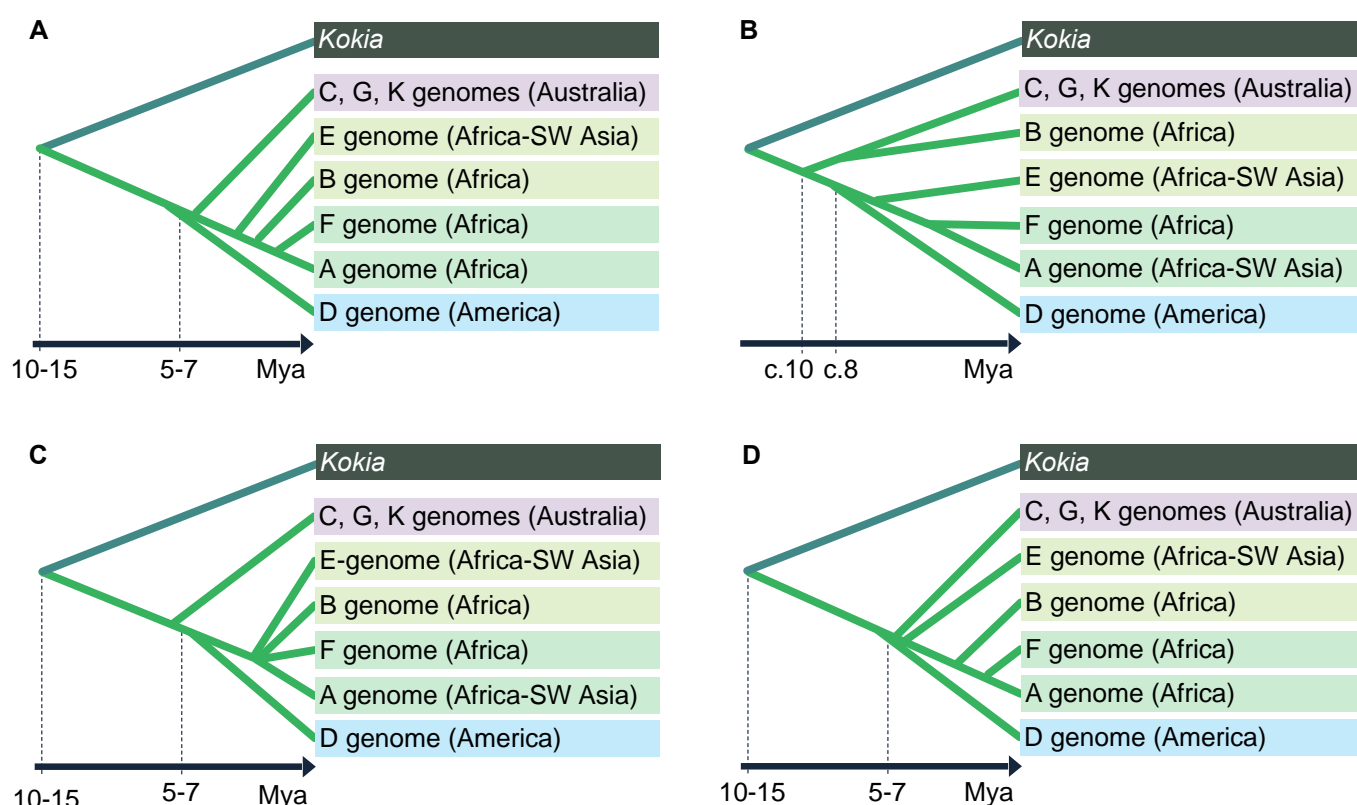


Figure 3. Trees comparing the main molecular phylogenetics-based hypotheses about the evolutionary history of *Gossypium* diploid genome groups. (A) D-genome clade diverging first (Cronn *et al.*, 2002). (B) C+G+K clade diverging first and B genome closer to C+G+K clade than to A+F+E clade (Chen *et al.*, 2016; Wu *et al.*, 2018). (C) C+G+K genomes clade diverging first (Wendel *et al.*, 2009). (D) D-genome clade diverging first and E-genome clade independent from the A+F+B clade (Hu *et al.*, 2019). Older estimated ages of divergence events in B relative to trees A, C and D (Chen *et al.*, 2016) resulted from a differing calibration of the molecular clock.

With respect to which of the trees in Figure 3 is “correct”, at present we can only point with confidence to one clear result, i.e., that there was an early and relatively rapid geographical as well as genomic diversification after the origin of the genus. This itself is an important biogeographical realization, irrespective of actual branching order at the base. Secondly, several other genome groups appear to have diversified in short temporal order, e.g. C+G+K, thus making modern day phylogenetic inference challenging. Given the availability of multiple genome sequences for all genome groups in *Gossypium*, it may be that a careful, genome-scale phylogenomics analysis will yield compelling insight into which of these hypotheses most probably reflects reality.

A repeated history of interspecific gene flow, chloroplast capture and long-distance hybridization

In addition to species and lineage *divergence*, the genus *Gossypium* appears to be veritably littered with examples of unexpected lineage *reunion*. Multiple studies have demonstrated that interspecific gene flow has been surprisingly common in the genus (Wendel *et al.*, 1991, 1995; Cronn *et al.*, 2003; Cronn and Wendel, 2004; Alvarez and Wendel, 2006), often between species that are geographically strongly disjunct and thus would seem to have no opportunity to hybridize. Some of these cases (Wendel and Cronn, 2003; Chen *et al.*, 2016; Wu *et al.*, 2018; Grover *et al.*, 2019) may represent the frequently reported phenomenon of “chloroplast capture” (Rieseberg and Soltis, 1991), which is thought to sometimes lead to cytonuclear incompatibility (Rieseberg and Soltis, 1991; Barnard-Kubow *et al.*, 2016), reproductive isolation and the origin of new lineages that subsequently appear to biologists as new species (Postel and Touzet, 2020; Zupok

et al., 2021). One infra-specific example of this for the American diploids concerns the Mexican species *G. aridum*, for which only populations from the Colima region contain plastid genomes that are more closely related to those found in the Baja California species than in other, non-Colima populations of *G. aridum* (Alvarez and Wendel, 2006).

More mysterious is the interspecific nuclear gene flow that has also been uncovered in *Gossypium* for the American diploid species *G. gossypioides*. This very narrowly distributed endemic from Oaxaca, Mexico is striking in its level of incongruence between phylogenetic data sets, as detailed in a number of studies (Wendel *et al.*, 1995; Cronn *et al.*, 2003). In addition to sharing a plastome with the Peruvian endemic *G. raimondii*, the nuclear genome of *G. gossypioides* was shown to contain African-like ribosomal DNAs and many other repetitive sequences as well (Wendel *et al.*, 1995; Zhao *et al.*, 1998; Cronn *et al.*, 2003). The processes and timing of these highly unusual introgression, recombination, and speciation events remain mysterious to this day.

In addition to these bizarre cases of inter-specific or even inter-genomic introgression that have been substantiated with molecular genetic analysis, data generated in the pre-molecular era also has bearing on the topic. For example, Fryxell (1971) used morphological characters to suggest a close relationship between the B-genome *G. triphyllum*, a species from southwestern Africa, and the Australian G-genome species (Fryxell, 1979)⁴. A possible explanation, at present untested, is that these similarities reflect hybridization of an ancestral B-genome plant with an immigrant Australian plant from the ancestor of the C and G genomes (Cronn *et al.*, 2002), with some genes from the Australian lineage now integrated into the modern African B-genome. Cytogenetic data also have been speculated to possibly implicate ancient hybridization in the origin of the E genome (Phillips, 1966; Cronn *et al.*, 2002), although this remains purely speculative at this time.

C - Phylogeny of American diploids, the subgenus *Houzingenia*

Most of the 13 D-genome diploid *Gossypium* species are endemic to western Mexico (Figure 4), revealing this region as the center of diversity for this clade (Wendel and Grover, 2015). These American diploid *Gossypium* species are suggested to have originated in northwestern Mexico, with later radiation to other regions (Wendel and Albert, 1992). Two species are endemic to northwestern South America, one from the Galapagos Islands, the other from northwestern Peru. Relationships among all of these species have been extensively studied, most recently using whole genome resequencing results (Grover *et al.*, 2019). A summary depiction of phylogenetic relationships is shown in Figure 4, where similarity of colors indicates phylogenetic proximity. A perusal of these colors reveals several more or less well-defined clades as well as some relatively recent long-distance dispersals involving, e.g., the close sister-species *G. davidsonii* from Baja California and *G. klotzschianum* from the Galapagos Islands.

These relationships become more evident when they are superimposed on the actual phylogeny (Figure 5), which shows species grouped according to taxonomic sections and subsections (Fryxell, 1979, 1992) and using the results from whole-genome phylogenomic analyses of Grover *et al.* (2019). While there are many areas of congruence between classical and modern views, there also are areas of incongruence. Some of these reflect the molecular-based discoveries on interspecific hybridization, as exemplified by *G. aridum*, *G. gossypioides* and *G. raimondii* discussed above. Other differences most likely are due to the deeper insights obtained from modern explicit phylogenetic approaches using massive datasets.

In all nuclear data generated to date, *G. gossypioides* appears as the branch earliest diverged from the main D-genome stem (Wendel and Cronn, 2003; Feng *et al.*, 2011; Yu-xiang *et al.*, 2013; Grover *et al.*, 2019). For now, *G. gossypioides* is taxonomically included in section *Eryoxylon*, as the only species of subsection *Selera* (Fryxell, 1979, 1992) but it clearly diverged (at ca. 2.6 MYA) well before the split (ca. 1.8 MYA) between sections *Eryoxylon* and *Houzingenia*. As discussed above, the cpDNA-based phylogenetic data indicates that the chloroplast genome of *G. gossypioides* appears most similar to that of *G. raimondii* (Wendel and Albert, 1992; Cronn *et al.*, 2003; Grover *et al.*, 2019), notably, the only species with which it is interfertile (Cronn *et al.*, 2003). Interestingly, the putatively African-like ribosomal and other repetitive sequences noted earlier (Wendel *et al.*, 1995; Zhao *et al.*, 1998; Cronn *et al.*, 2003) in the genome of *G.*

⁴ p. 53 in Fryxell (1979)

gossypioides were not detected in this most recent study (Grover *et al.*, 2019). *Gossypium gossypioides* thus shows strangely contradictory phylogenetic affinities whether morphological (Fryxell, 1979), cytogenetic (Brown and Menzel, 1952), nuclear genome or chloroplast genome (Grover *et al.*, 2019) characters are studied.

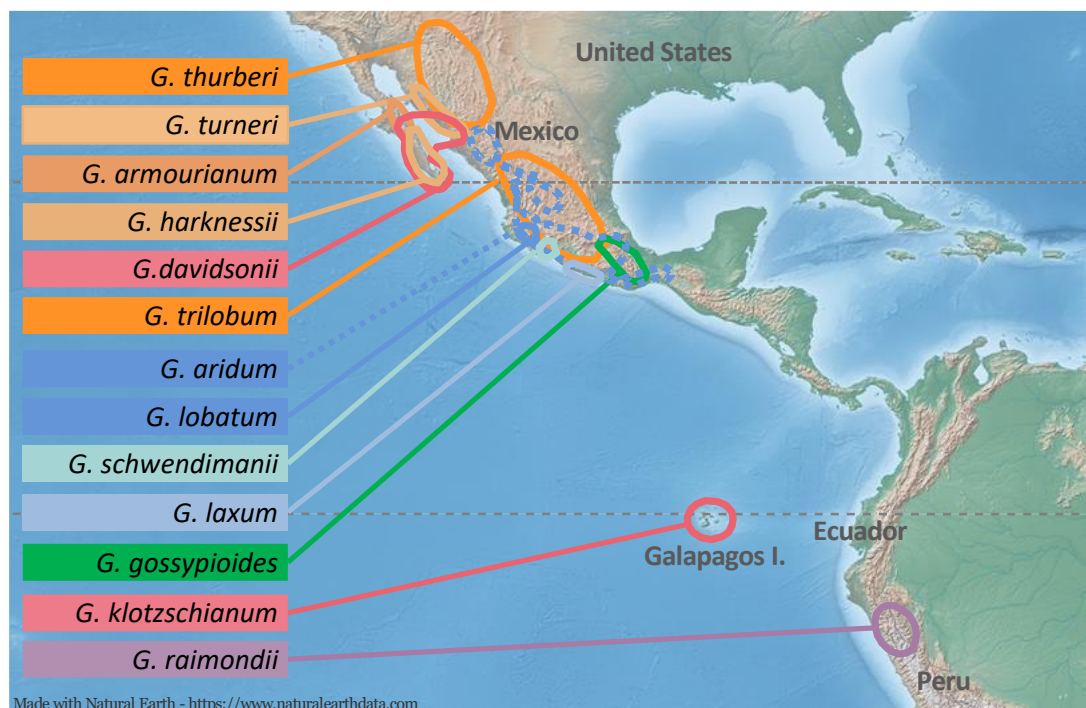


Figure 4. Geographic distribution of the D-genome diploid *Gossypium* species. Sources: (Ulloa *et al.*, 2006; Ulloa, 2014; Grover *et al.*, 2019; Jenderek and Frelichowski, 2019).

What might be the explanation for the unusually high level of phylogenetic conflict for *G. gossypioides*? First, the close affinity of its cpDNA with that of *G. raimondii*, and the presence of A-genome sequences in its nuclear genome, led Wendel *et al.* (1995) to hypothesize that this lineage could have arisen as a diploid offshoot of the hybridization event that ultimately generated the AD-genome allopolyploids (these discussed in the next section); as summarized below, *G. raimondii* is the species whose genome is genetically closest to the allopolyploid subgenome Dt, and its nuclear genome is very different from that of *G. gossypioides* (Fig. 5). Thus, the “diploid derivative from initial A x D hybridization” hypothesis fails to account for how different the two diploid D-genome genomes are. As an alternative, perhaps the similarity of the chloroplast genome of *G. gossypioides* to that of *G. raimondii* is explained by a natural hybridization and chloroplast capture event (from the latter into the former species) prior to the dispersal of the ancestor of the latter species to S. America. This scenario implies a close geographic proximity of plants from the ancestral lineages of *G. raimondii* and *G. gossypioides* at the moment of the hypothetical hybridization.

A second species for which there is conflict between the results of modern phylogenetics with the views of classical taxonomy is *G. raimondii*. As shown in Figure 5, this species is cladistically linked by all molecular data (Grover *et al.*, 2019) with species of subsection *Integrifolia* in section *Houzingenia*. Taxonomically (Fryxell, 1979, 1992) however, *G. raimondii* was placed in section *Eryoxylon*. We note that *G. raimondii* had previously been considered as a subspecies of *G. klotzschianum*, i.e., *G. klotzschianum* ssp. *raimondii* (Fryxell, 1979), which is interesting in that *G. klotzschianum* is one of the two other species of subsection *Integrifolia* to which *G. raimondii* is linked in phylogenetic studies (Figure 5).

A third D-genome species for which phylogenetic data yielded unexpected results is *G. aridum*. As noted above, plants in part of its wide geographical range, in the Mexican state of Colima, share a chloroplast genome with the Baja California–Galapagos Islands species pair *G. davidsonii* and *G. klotzschianum* (DeJooode and Wendel, 1992; Alvarez and Wendel, 2006; Grover *et al.*, 2019) while all other *G. aridum* populations show a chloroplast genome phylogenetically close to the other species of the subsection

Eryoxylon to which *G. aridum* belongs. A cytoplasmic introgression is the main hypothesis to explain the differing chloroplast genome in the Colima population (Grover *et al.*, 2019).

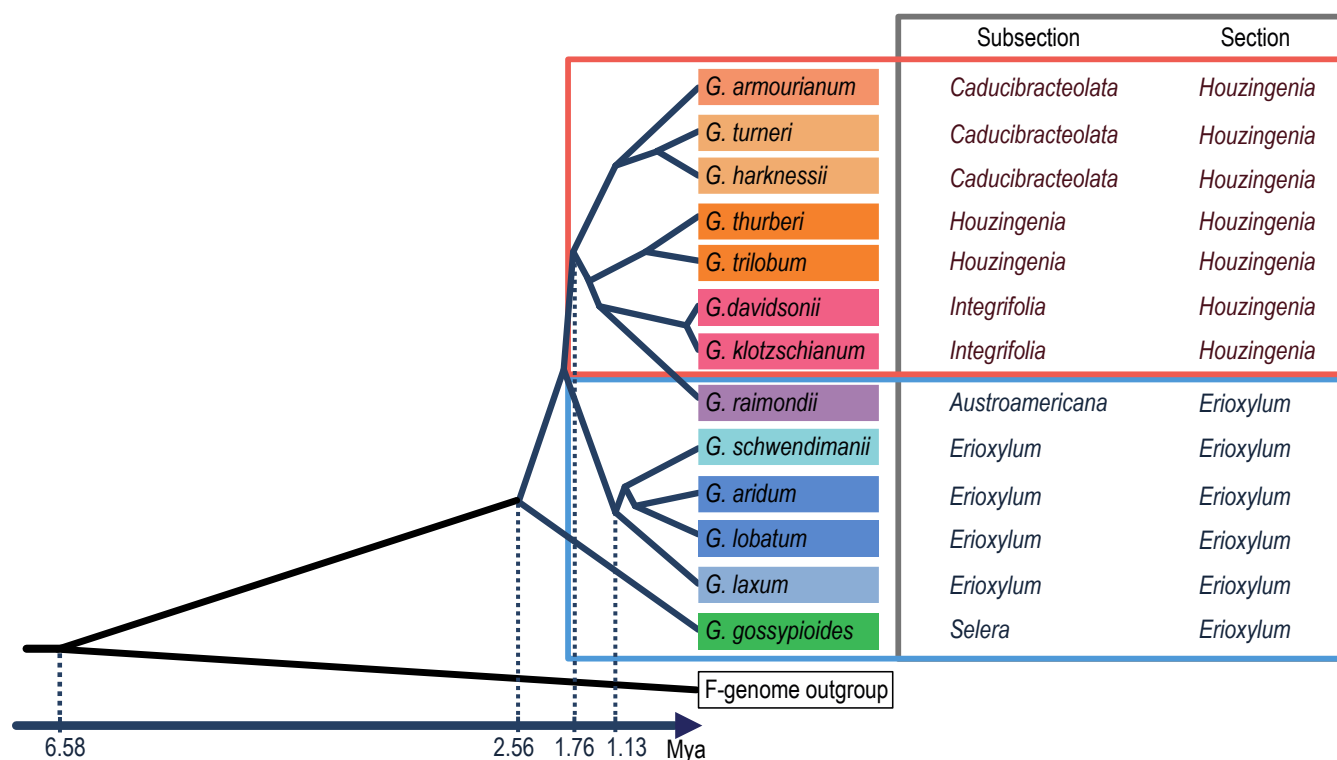


Figure 5. Correspondence of classical taxonomy and nuclear genes-based phylogenetic relationships of D-genome species. Sources: Fryxell, 1979, 1992; Grover *et al.*, 2019. Time scale: million years ago.

As shown in Figure 4, the two taxonomic sections of D-genome species exhibit rather different ranges in geographic distribution; species of section *Houzingenia* (including *G. raimondii* inside *Houzingenia*) are rather widely distributed, from northwestern Mexico to the Galapagos Islands and Peru, while species of section *Erioxylum* occupy rather small areas - except for *G. aridum* - along the Pacific coast and western cordillera ranges of central and southern Mexico. Section *Houzingenia* is also older - 1.76 MYA - than section *Erioxylum* - 1.13 MYA - (Grover *et al.*, 2019), which perhaps is related to the aggregate differences in range. The hypothesis that American D-genome diploids initially emerged in northwestern Mexico, and later radiated to other regions (Wendel and Albert, 1992), is in agreement with the molecular genetics-based ages of these two sections. This biogeographic speculation also is of interest for its relevance to the allopolyploid clade, in that its origin involves a species from section *Houzingenia*, more precisely, the ancestor of modern *G. raimondii*. This is discussed in the following section.

D - Allopolyploidization and evolution of the allopolyploid clade, subgenus Karpas

The allopolyploid *Gossypium* species originated through a natural hybridization between an A-genome-like plant and a D-genome-like plant followed by chromosome doubling (Skovsted, 1934; Beasley, 1940; Senchina *et al.*, 2003; Wendel and Grover, 2015; Hu *et al.*, 2021). The resulting polyploid genome, thus composed of two subgenomes, named At and Dt, shows an amphidiploid behavior, that is, similar to that of a diploid (Brown, 1951). This diploid-like behavior may be related to the two-fold difference in size between the chromosomes of the A and D genomes (see Table 1), and/or to other factors that are subject to strong positive selection in nascent allopolyploids that enforce homologous as opposed to homoeologous chromosome pairing. The cytoplasmic organelles of the allopolyploid lineage, i.e., those of the chloroplasts and mitochondria, are genetically close to the organelles of the A-genome lineage (Wendel, 1989; Chen *et al.*, 2017); these cytoplasmic organelles are maternally transmitted in *Gossypium*, and thus it is clear that the A-genome progenitor of the allopolyploids served as the female, or seed parent.

Following decades of speculation and study concerning the best modern models of the A and D genome donors in the initial allopolyploidization event (reviewed in Endrizzi *et al.* (1985), Wendel and Grover (2015)), it now is clear that the sister-species *G. herbaceum* and *G. arboreum*, the only two extant A-genome species, appear equally good as models of the ancestor of the allopolyploid At subgenome. There now also is convincing evidence that *G. raimondii* is the best living model of the ancestor of the Dt subgenome. Two major reciprocal translocations, between chromosomes A02 and A03 and between chromosomes A04 and A05, are common to the allopolyploid *Gossypium* species; as these translocations are not observed in the A-genome species, they are hypothesized as contemporaneous or posterior to the polyploidization, and their occurrence in all of the extant allopolyploid cotton species makes most probable a monophyletic origin for the allopolyploid lineage (Menzel and Brown, 1954; Wendel and Cronn, 2003; Hu *et al.*, 2019; Huang *et al.*, 2020). The monophyly of the allopolyploid cotton lineage was phylogenetically tested and confirmed using genome sequence data (Grover *et al.*, 2012).

1. Age of the *Gossypium* allopolyploid lineage

One of the early successes of the molecular era was the advent of the molecular clock, which for the first time provided the opportunity to date the A vs D genome divergence and later allopolyploidization (Wendel, 1989). The initial results were definitive in excluding the many speculations regarding the timing of allopolyploid formation, which ranged from early Cretaceous to modern, the latter meaning human-created in agricultural or pre-agricultural times. Most satisfying was that the estimated time of initial allopolyploidization, 1 to 2 MYA (Wendel, 1989; Senchina *et al.*, 2003; Li *et al.*, 2015), was consistent with some of the earlier suggestions based on cytogenetic studies and ecological-evolutionary speculations (Phillips, 1963; Fryxell, 1979). All data subsequently generated, including those based on entire genome sequences, are consistent with this indication of a mid-Pleistocene origin for allopolyploid *Gossypium*, although the precise estimates vary somewhat based on the rate estimates used for clock dating.

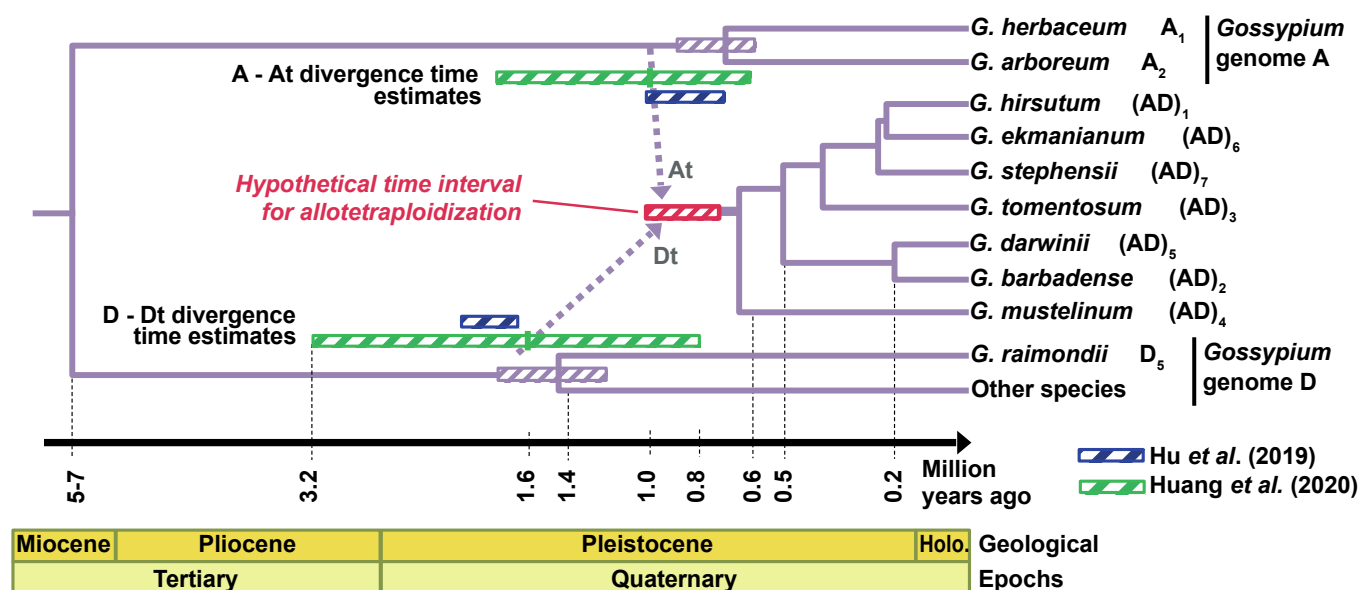


Figure 6. Temporal depiction of important events in *Gossypium* allopolyploidization as deduced from molecular clock-based dating. Logarithmic time scale in million years. Sources for time ranges: Grover *et al.*, 2019; Hu *et al.*, 2019; Chen *et al.*, 2020; Huang *et al.*, 2020.

Importantly, the cpDNA based data (Wendel, 1989; Chen *et al.*, 2016) represent the divergence between modern A-genome species and the At genome of the allopolyploid, whereas nuDNA data can date both the A vs D divergence as well as the At vs. Dt divergence; these, in fact, yield comparable and consistent results (Senchina *et al.*, 2003; Page *et al.*, 2013; Chen *et al.*, 2020). Because we aren't able to use the *actual* genome donors, but only their best *models*, the molecular clock-based estimates are by definition, upper bounds, or maximum ages for the allopolyploidization event (Fig. 6). An additional result that is confirmed by multiple studies is that the subgenome Dt of the allopolyploids is less similar to D-genome diploid *G.*

raimondii than the subgenome At of the allopolyploids is to the genomes of A-genome diploids *G. herbaceum* and *G. arboreum* (Senchina *et al.*, 2003; Page *et al.*, 2013; Chen *et al.*, 2020).

Estimated time ranges (Fig. 6) based on this merger of phylogenetics with molecular clock dating permit several conclusions. First, the two A-genome species diverged from each other after the origin of the allopolyploids, perhaps ca. 0.7 (0.89–0.61) MYA (Huang *et al.*, 2020). Second, the D-genome donor, with the best model being *G. raimondii*, is about 50% more divergent from genome Dt as the A-genome species are from genome At (Chen *et al.*, 2020). Third, allopolyploidization likely occurred about 0.7 to 1.0 MYA (red bar, Fig. 6). These results lead to a fourth important conclusion, namely, that the first divergence in the allopolyploid clade, represented by modern *G. mustelinum*, didn't occur until about 0.63 MYA (Chen *et al.*, 2020). Accordingly, no information exists on possible earlier allopolyploid lineages, i.e., those that may have arisen but subsequently went extinct during the first 100,000–300,000 years of allopolyploid existence.

The divergence time of *G. darwinii* and *G. barbadense* was estimated at ca. 0.2 MYA on the basis of genomic data (Wu *et al.*, 2018; Chen *et al.*, 2020), indicating neatly distinct species. Morphologically, the two species have been described as very close, with *G. darwinii* once being taxonomically considered a subspecies of *G. barbadense* and even discussed as a possible wild ancestor for domesticated *G. barbadense* (Fryxell, 1979). Part of *G. darwinii*'s diversity, however, is best explained by an interspecific gene flow from introduced, *G. hirsutum*-introgressed, *G. barbadense* (Wendel and Percy, 1990).

2. Location of the allopolyploidization event

Various hypotheses have been forwarded about the location of the allopolyploidization that generated the present-day *Gossypium* polyploid clade. These include northwestern South America, particularly the northwestern coast of Peru close to *G. raimondii* and *G. barbadense* wild populations (Hutchinson *et al.*, 1947; Wendel and Grover, 2015; Chen *et al.*, 2020); northeastern Brazil close to the area occupied by *G. mustelinum* (Ano *et al.*, 1982); northwestern Mexico (Hancock, 2012)⁵; and Mesoamerica, in the region of the Isthmus of Tehuantepec close to the geographic area occupied by *G. gossypoides* (Wendel *et al.*, 1995, 2010; Liu *et al.*, 2001; Grover *et al.*, 2012). Adding to the uncertainty is the fact that there is no evidence of there ever having been any natural A-genome population in the Americas, and the geographic peculiarity that the closest model of the D-genome ancestor, *G. raimondii*, is from northwestern Peru. Wild populations of AD-genome allopolyploid *Gossypium* species are widely dispersed over Mesoamerica, including one small region of the Dominican Republic in Hispaniola, northwestern South America, northeastern Brazil, and even vastly dispersed populations in the Pacific, including Hawaii and the Wake Atoll (see Figure 1 and 7). Thus, it is difficult to make strong assertions about the location of the initial hybridization event that gave rise to the allopolyploids.

These uncertainties notwithstanding, it is reasonable to suggest that the A-genome female x D-genome male hybridization occurred in the region extending from Mesoamerica to northern South America, because this is where extant D-genome species are located. Within these bounds, perhaps the least complex hypothesis might be that hybridization occurred in the region near modern *G. raimondii* populations in northwestern South America, followed much later by dispersal events accompanying the allopolyploid radiation, as shown in Figure 7. The hypothesis of an initial allopolyploid in northeastern Brazil (Ano *et al.*, 1982) has some support, if only tentative, in that this is the area occupied by *G. mustelinum*, representing one branch of the earliest-divergence among allopolyploid taxa. Also, it is the point of the Americas closest to Africa and hence a source population of an A-genome ancestor that acted as the female parent. Dispersal of the allopolyploids from this region would have followed the prevailing marine currents along the eastern coast of South America towards the Caribbean (Ano *et al.*, 1982). One challenge to this scenario is that there is no evidence for any D-genome plant from the lineage leading to *G. raimondii* from northeastern Brazil, or from anywhere east of the Andes or central cordillera of Mexico and Central America.

⁵ Hancock (2012) p.185

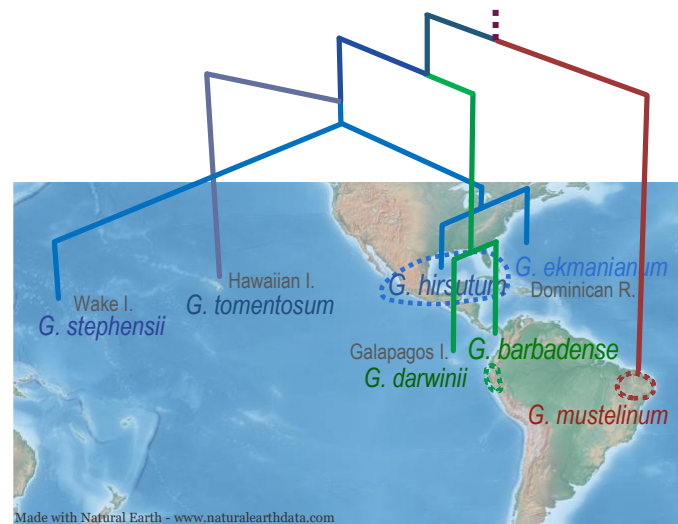


Figure 7. Phylogenetic diversification and initial geographical dispersion of the seven extant allopolyploid *Gossypium* species, starting from a hypothetical location in northwestern Peru, or alternatively, western Mexico, of the allopolyploidization event. Source: Grover *et al.* (2019).

Another possibility is that allopolyploidization occurred in the region presently occupied by *G. gossypoides*, near the isthmus of Tehuantepec in Mesoamerica, or perhaps anywhere from there northward into the center of diversity of the D-genome clade in western Mexico. The presence of the closest D-genome model *G. raimondii* in Peru would then have to be postulated to reflect a more recent long-distance dispersal event to that region; given that this phenomenon is a primary biogeographic feature of the evolution of *Gossypium*, and that this dispersal event may have been more recent than the origin of allopolyploidy in the genus, this certainly is a reasonable hypothesis.

3. Diversification and geographical dispersal of the allopolyploid clade

As introduced in Figure 2, phylogenetic data demonstrate three primary clades of allopolyploid cottons. One branch of the earliest split is occupied exclusively by *G. mustelinum*, with the remaining 6 species separated into two more recently diverged sister groups, the *G. barbadense*-*G. darwinii* clade and the *G. tomentosum*-*G. hirsutum*-*G. ekmanianum*-*G. stephensii* clade (see in Fig. 8 examples of cotton plants of wild allopolyploid species in real environment). This partition in three clades is now firmly established by a wealth of phylogenomic data (Wendel and Grover, 2015; Gallagher *et al.*, 2017; Hu *et al.*, 2021) and finds additional support from chromosomal inversions (Wang *et al.*, 2016).

A notable feature of allopolyploid cottons, compared to diploid cottons, has been emphasized by Fryxell (1979) among others: unlike most diploid species, allopolyploids, in their wild forms, occupy coastal habitats, excepting *G. mustelinum* in its semi-arid habitat in northeastern Brazil (da Silva *et al.*, 2018). It has been suggested that adaptation to coastal environments was enabled as an emergent and novel property of allopolyploidization (Fryxell, 1979; Wendel and Grover, 2015), although we also note that the species of the section *Houzingenia* to which the D-genome progenitor belongs show a similar but less marked behavior. *Gossypium* seeds are tolerant to prolonged stays in salty waters, which, along with their origin during the large fluctuations in ocean levels during the Pleistocene, promoted the dynamic of geographical dispersion and evolutionary radiation (Fryxell, 1979). As mentioned earlier, many of the allopolyploid species are endemic to island environments, *G. darwinii*, *G. tomentosum*, *G. ekmanianum* and *G. stephensii*, and wild forms of *G. hirsutum* almost always occur in dry coastal areas or far-flung islands in the Pacific (Fryxell, 1979; Stephens, 1958).



Figure 8. **a.** *G. hirsutum* race *yucatanense*, old plant in a coastal prairie of *Batis maritima* routinely inundated by salt water during storms. Everglades National Park, Florida, United States; **b.** *Gossypium tomentosum*, Kahoolawe, Hawaii; **c.** *Gossypium darwinii*, Parque Nacional, San Cristóbal Island, Galápagos; **d.** *Gossypium darwinii*, sandy beach of Puerto Grande, San Cristóbal Island, Galápagos; **e.** *Gossypium stephensii*, Wake Atoll.

III - Domestication of the two cultivated allopolyploid cotton species

Of the seven allopolyploid species, only two have been domesticated by humans. Plant domestication can be defined as genetic adaptation to cultivation over many generations under the more or less conscious and directed influence of humans (Zeder, 2015). For cotton, the domestication syndrome includes multiple traits, including epidermal seed trichomes, the “fiber”, that is more easily detachable from the seed, more abundant fiber through bigger bolls and seeds, longer fibers, and readily germinating seeds thanks to a thinner, more permeable seed coat and reduced dormancy (Fryxell, 1979; Brubaker *et al.*, 1999). Traits that were later selected, over millennia of crop improvement, are supposed to include fineness and length of fiber, white fiber, a more compact plant architecture permitting both greater synchrony in flowering and a shorter life cycle as well as easier field management, and loss of photoperiod sensitivity from short-day flowering to daylength neutral so that the plants no longer depend on changes in day-length to induce flowering (Gross and Strasburg, 2010). Shorter and coarser fibers more firmly attached to the seed coat reliably characterize truly wild cotton plants (Hutchinson, 1951); these wild plants also show ca. 10 to 12% fiber as a percent of the seed plus fiber weight, while it usually is above 20% for the domesticated germplasm (Fryxell, 1979). In practice, however, there is a full continuum in fiber weight percentage, for many reasons but perhaps most commonly because of escape from cultivation at various stages over thousands of years, and reestablishment as feral derivatives in native vegetation. As for archaeological remains, most often only a few traits characteristic of the domestication syndrome can be evaluated: the size and morphology of the seeds, including seed diameter and length that permit a calculation of 'seed grade' that is highly correlated with seed weight; boll diameter and length, the latter used to calculate a 'boll index'; and boll size; boll locule (carpel)

number; and fiber length. The length, fineness and whiteness of the lint can also be studied in yarns or fabrics. Statistically, wild or more primitive plants have smaller seeds, fewer locules per boll, smaller bolls overall and a shorter, dingy tan-brown to off-white lint (Smith and Stephens, 1971; Wendel, personal observations).

A - Archaeological data on textile use of cotton in Mesoamerica and South America

1. Mesoamerica and *G. hirsutum* cotton archaeology

In Mesoamerica, the humid and hot environment prevailing in many regions causes a rather rapid degradation of organic materials (Filloy Nadal, 2017) and remains of plant fiber textiles from thousands of years ago are in general very rare. In Figure 8 are shown the locations and ages of the earliest archaeological cotton remains, proving a textile use or cultivation of cotton in Mesoamerica, the hypothetical centers for the initial domestication of *G. hirsutum* and the primary sites related to the domestication of other crops.

The textile use of cotton in Mesoamerica is documented by the end of the Early Formative period (4,000-3,000 YBP), when cotton appears in textile manufacturing and loom weaving in archaeological sites of the Tehuacan Valley and the Valley of Mexico. In the Tehuacan Valley caves remains of cotton fiber dated to ca. 5,500 years old were found, as were parts of cotton bolls perhaps of comparable age (without radiocarbon dating (Wendel *et al.*, 1992)). Cotton is considered cultivated during the "Ajalpan" archaeological phase (3,500-2,900 YBP) of Tehuacan Valley and cotton remains appear consistently, along with cloth woven from cotton fiber, during the "Santa Maria" phase (2,900-2,200 YBP) of this archaeological site (MacNeish *et al.*, 1967). Some hint of cotton cultivation was also found at ca. 3,100 YBP in Belize (Miksicek, 1990); archaeological *Gossypium* pollen in the Gulf Coast of Tabasco appears at 4,500 YBP with no proven link to cotton cultivation (Pope *et al.*, 2001). All of the archaeological remains of cotton bolls and seeds found in Mesoamerica and Central America were characterized as belonging to the species *G. hirsutum*.

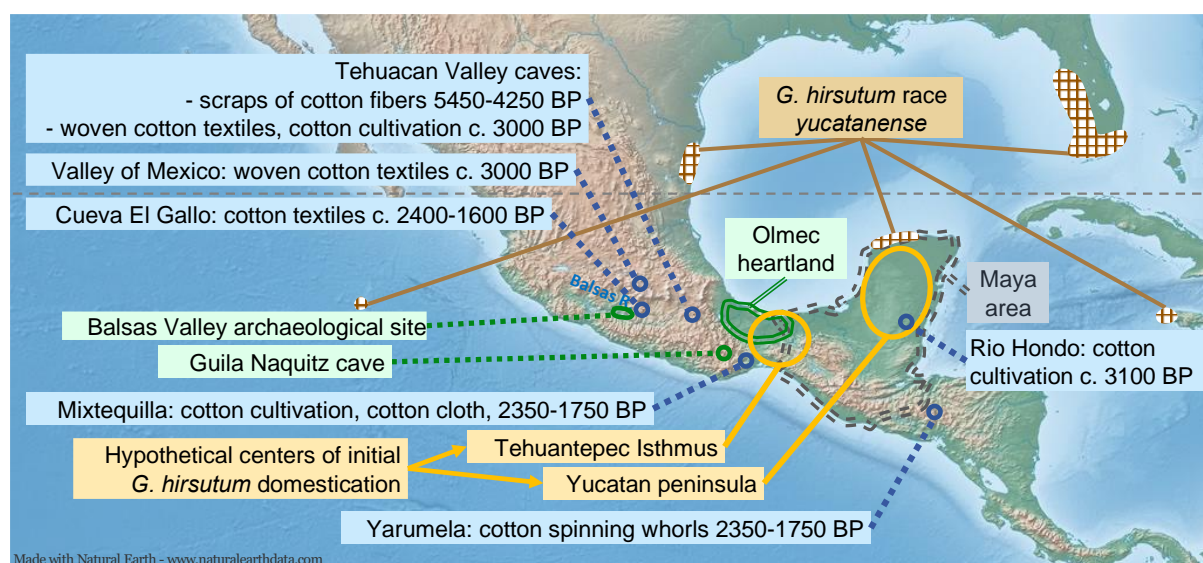


Figure 9. Location of the earliest archaeological cotton remains in Mesoamerica, of hypothetical centers of initial domestication of *G. hirsutum* and of main sites related to crop domestication in these same regions. Sources: MacNeish, 1967b; MacNeish *et al.*, 1967; Smith, 1967; Stephens, 1973; Miksicek, 1990; Lentz *et al.*, 1997; Follensbee, 2008; Filloy Nadal, 2017; Grover *et al.*, 2019.

The sites of Guilá Naquitz cave and Balsas Valley featured in Figure 9 correspond to the probable initial domestication centers of squash, beans and corn - the *Three Sisters* of North American Indigenous agriculture - and other crops, beginning ca. 10,000 years YBP and spanning many millennia (Smith, 1997). The oldest known civilization in Mesoamerica is that of the Olmec people, which flourished from roughly 3,400 to 2,400 YBP, while Pre-Olmec and early Olmec cultures existed from about 4,500 to 3,400 YBP (Pool, 2007). The Olmec area (see Figure 9) in the region of Tres Zapotes, San Lorenzo and La Venta, close to the Yucatan peninsula, is characterized by swampy lowlands, low hills and ridges and a warm humid climate unfavorable to a wild cotton population and to cotton cultivation (Smith and Stephens, 1971), and also to textile preservation. The Olmec civilization begins to show archaeological remains of cotton spinning

implements and cotton textiles only in the second half of its existence, between 3,000 to 2,400 YBP (Follensbee, 2008).

The Maya civilization developed over a Preclassic period, 3,750 to 1,750 YBP, and a Classic period, 1,750 to 1,050 YBP, and geographically from modern-day Guatemala and Belize to southern Mexico and the entire Yucatan peninsula (Estrada-Belli, 2010). An ancestor of all present-day Maya languages, Proto-Mayan, is estimated to have existed before 4,200 YBP (Campbell, 1997); older still, some Maya settlements in Guatemala are thought to be older than 6,000 YBP (Söchtig *et al.*, 2015).

2. South America and *G. barbadense* cotton archaeology

In South America, archaeological evidence attests to a very old textile use of plant fibers, such as bark-fiber-like coils and complex vegetal textiles from *Agavaceae*, *Bromeliaceae* and *Cyperaceae* found in Guitarrero Cave, high in the Andes Mountains in Peru, dated to the 12th millennium BP (Dillehay *et al.*, 2007; Jolie *et al.*, 2011). In Monte Verde, on the Pacific coast of southern Chile, plant-derived cordage and cordage impressions have been dated to c.14,500 YBP (Dillehay *et al.*, 2008). Figure 10 indicates the location and estimated ages of evidence about cotton use, cultivation and domestication along the South American Pacific coast and Western slopes of the Andes.

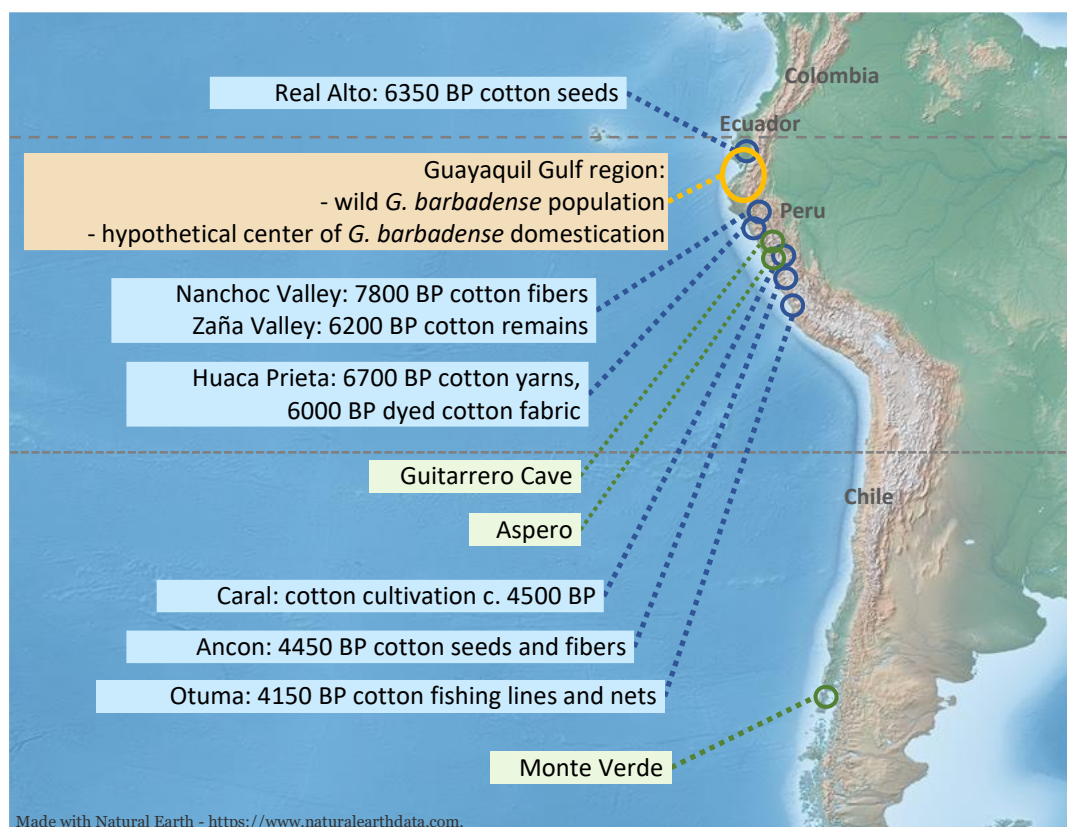


Figure 10. Location of the earliest archaeological cotton remains in South America, some main contemporaneous archaeological sites along the Pacific coast and the presumed wild range of *G. barbadense* and its hypothetical domestication center. Sources: Stephens and Moseley, 1974; Damp and Pearsall, 1994; Dillehay *et al.*, 2007, 2012; Piperno, 2011; Splitstoser *et al.*, 2016; Beresford-Jones *et al.*, 2018.

Cotton remains associated with human settlements in South America recovered in coastal central and northern Peru and coastal Ecuador were dated to the beginning of the 8th millennium BP (Splitstoser *et al.*, 2016), that is, as old as the oldest evidence of cotton fiber use in the Old World (Moulherat *et al.*, 2002). Cotton cultivation might have begun in the middle of the 7th millennium BP; the species was characterized as being *G. barbadense* whenever specific status could be assessed. The oldest elaborated cotton material is yarn dated to 6,882–6,657 years YBP found at Huaca Prieta (Dillehay *et al.*, 2012). Also in this site, one impressive find is a 6,000-year-old dyed cotton fabric (Splitstoser *et al.*, 2016); it is the earliest recorded use

of indigo dye to date worldwide. In Huaca Prieta, first human presence is dated to between 13,720 and 13,260 YBP and was represented by maritime foragers (Dillehay *et al.*, 2012). Among the other archaeological sites featured in Figure 10, Aspero is a 5,650 to 4,450 year old fishing settlement at the mouth of the Supe river, where fishing nets and lines were made from cotton grown with irrigation in the inland site of Caral (Beresford-Jones *et al.*, 2018). The ages of archaeological cotton remains of South America contrast with the much younger comparable remains found in Mesoamerica. Nevertheless, the unfavorable conditions for long-time preservation of organic matter in most habitats of Mesoamerica probably introduce a bias, as a much higher proportion of the older textile remains must have disappeared than in the dry environments of the western slopes and plateaus of the Andes.

B – Taxonomy of the two domesticated allopolyploids

The word *Gossypium* was used in the first edition of *Species Plantarum* (Linnaeus, 1753) for the binomial taxonomic system that the book first developed, as the genus name for the cultivated cotton plants. The scientific naming of the two cultivated New World cotton plants, *G. hirsutum* and *G. barbadense*, are also Linnaean in origin, followed about two hundred years of European exploration and collection in the Americas. American cotton plants were first described by travelers and botanists in the 16th century (see, e.g., de Lery, 1580). Because of the wide range of morphology observed among populations from different regions, reflecting the vastness of the geographic ranges of cultivation, the variety in the stages of domestication reflected in these populations, and interspecific hybridization, a large number of “species” were described by subsequent scientists (Table 2). Paul Fryxell in particular has written extensively on this topic (Fryxell, 1979), even publishing a small book devoted this subject (Fryxell, 1976), but the main point is that these many “species” were ultimately synonymized to reflect the biological reality that this morass of biological diversity in fact encompasses only two species, *G. hirsutum* L. and *G. barbadense* L.

Table 2. Synonymy *Gossypium hirsutum* L. and *G. barbadense* L. Sources: Bauhin, 1623; Linnaeus, 1753, 1763; Fryxell, 1979; Grover *et al.*, 2019.

	Pre-Linnaean scientific naming (Bauhin, 1623)	Linnaean scientific naming (Fryxell, 1979)
<i>Gossypium hirsutum</i> L.	<i>Gossipium frutescens semine albo</i>	<i>Gossypium hirsutum</i> L., 1753 <i>Gossypium religiosum</i> L., 1767 <i>Gossypium latifolium</i> Murr., 1776 <i>Gossypium punctatum</i> Schumach. & Thonn., 1827 <i>Gossypium mexicanum</i> Tod., 1868 <i>Gossypium marie-galante</i> Watt, 1927
<i>Gossypium barbadense</i> L.	<i>Gossipium Brasilianum</i> , Bellon. & Belli	<i>Gossypium barbadense</i> L., 1753 <i>Gossypium vitifolium</i> Lam., 1786 <i>Gossypium peruvianum</i> Cav., 1788 <i>Gossypium brasiliense</i> Macfad., 1837 <i>Gossypium lapideum</i> Tussac, 1818 <i>Gossypium barbadense</i> var. <i>integrum</i> Griseb., 1859 <i>Hibiscus barbadensis</i> (L.) Kuntze, 1891 <i>Gossypium barbadense</i> f. <i>hahnii</i> Roberty, 1950 <i>Gossypium barbadense</i> f. <i>eggersii</i> Roberty, 1950

C - Domestication of Gossypium barbadense

1. Timeline and location of G. barbadense domestication

The synthesis of archaeological studies (see above) indicates that *G. barbadense* probably was the species used first for textiles among American cotton species. Wild *G. barbadense* plants are relatively rare, but still likely occur in the region surrounding the Guayaquil Gulf in southwest Ecuador and northwestern Peru, along with the more common dooryard and feral arborescent plants (Percy and Wendel, 1990; Damp and Pearsall, 1994; Westengen *et al.*, 2005). Archaeological plant remains found on the coast and in valleys of this region show primitive morphological characteristics, small “bolls” (colloquial term for the botanically

correct "capsules") and more or less brownish lint, indicating they corresponded to early domesticated forms, from which it was concluded that original domestication was initiated in sites around there (Hutchinson *et al.*, 1947; Stephens, 1973; Westengen *et al.*, 2005; Wendel *et al.*, 2010). Molecular data (Percy and Wendel, 1990; Westengen *et al.*, 2005) revealed a center of genetic diversity that is geographically congruent with this hypothesis; DNA data particularly indicate a high diversity in the coastal departments of Peru.

The sites in South America showing the earliest proof of cotton cultivation lie on the Pacific coast from central Ecuador (Real Alto) to northeastern Peru (Huaca Prieta, Nanchoc and Zaña valleys). In Real Alto ca. 6,350 YBP cotton remains were considered clear signs of cotton domestication (Damp and Pearsall, 1994). In Huaca Prieta and the valleys of Nanchoc and Zaña, elaborated cotton textiles were dated to 6,700 to 6,000 YBP and cotton cultivation there can be assumed as these places are unfavorable to spontaneous cotton populations (Beresford-Jones *et al.*, 2018). The textile use of *G. barbadense* cotton for fishing nets, bags and diverse materials appears to have become widespread from Ecuador to northern Peru by 6,000 to 5,000 YBP (Splitstoser *et al.*, 2016; Beresford-Jones *et al.*, 2018).

Evidence thus seems to indicate that cotton was cultivated by the middle of the seventh millennium YBP; it can reasonably be assumed that the domestication of cotton had begun some time before. The region close to the Gulf of Guayaquil is, as explained above, considered the place where wild *G. barbadense* plants could initially entered into the domestication process; the closest archaeological sites possibly involved are Real Alto and the Nanchoc and Zaña valleys. From the time of the first interest in cotton, dated at ca. 7,800 YBP in Nanchoc valley, to hypothetical cotton cultivation, about 6,500 YBP in diverse places, some of the expected morphological changes associated with directional selection imposed by intentional domestication begin to be evident. Seeds, fiber samples, and boll parts dated from approximately 4,500 to 3,700 YBP found at sites in the Ancon-Chillon area of Peru resembled closely those still to be found in present-day cultivars in coastal Peru and Ecuador, with however some traits - small beaked bolls, fuzzy seeds, narrow fiber diameter - closer to the wild forms (Stephens and Moseley, 1974). Domestication and crop improvement are continuous processes, non-linear in their progress and not at equivalent rates for all traits.

A possible impetus for the early development of cotton cultivation in northwestern South America between ca. 6,000 and 4,000 YBP might have been fishing and the associated fishing nets (Moseley, 1992). Archaeological fishing sites close to the Pacific shoreline appears also to have been associated with settlements more inland that were dedicated to irrigated agriculture, including cotton cultivation. The main example is Aspero on the coast with Caral 20 km inland. Cotton from this region shows excellent technical properties for fishing lines and nets, such as resistance, durability and fineness of the threads. Fibers extracted from milkweed (*Asclepias* spp.) or dogbane (*Apocynum* spp.) are not, in fact, much inferior for these uses but these bast fibers need a tedious and long preparation (Beresford-Jones *et al.*, 2018) while cotton lint is readily used after its harvest and the separation of fibers from seeds. Cotton cultivation, also, could derive directly from the seeds separated from the fibers that were falling close to human dwellings, initiating a domestication scenario through door-yard agriculture (Lubbers and Chee, 2009). As perennial shrubs in direct proximity to human habitation, cotton plants would have provided a secure production of fiber; milkweed and dogbane probably were not cultivated and, taken from the local vegetation, could locally have become less frequent (Beresford-Jones *et al.*, 2018).

While *G. barbadense* cotton domestication most likely began in the Guayaquil Gulf region, a domestication center East of the Andes was also hypothesized to explain the Amazonian type of *G. barbadense* (Stephens, 1973; Phillips, 1976). Archaeological evidence and molecular data, however, do not support this hypothesis (Percy and Wendel, 1990; Lazo, 1991).

2. Intraspecific taxa in *G. barbadense*

Prior to European colonization, *G. barbadense* already had spread over a vast geographic area in South America, Central America and the Antilles as a result of human dispersal of more or less primitive domesticated forms (Lazo, 1991). The primary dispersals of *G. barbadense* from its domestication center in the Guayaquil Gulf region are hypothesized to have been a very early expansion to coastal southern Peru, a trans-Andean expansion before 2,500 YBP into the Amazon basin through the Marañon River valley and into northern South America through the Patia river and the Cauca river valleys (Lazo, 1991), and then into the Caribbean, Central America and southern South America (Figure 11). These dispersals were

accompanied by a diversification (var. *brasiliense*, type Sea Island) and by sequential reductions of genetic diversity and allelic richness (Percy and Wendel, 1990). The original west-Andean domesticated form is named type Ecuadorian. A close affinity was found between accessions from east of the Andes, the Caribbean, and Central America (Percy and Wendel, 1990), regions which thus globally define the area of *G. barbadense* var. *brasiliense*. The type Sea Island originated by a dispersal from a Peruvian-Ecuadorian stock different from the dispersal that gave rise to *G. barbadense* var. *brasiliense*, as shown by molecular data (Percy and Wendel, 1990; Ding *et al.*, 2015).

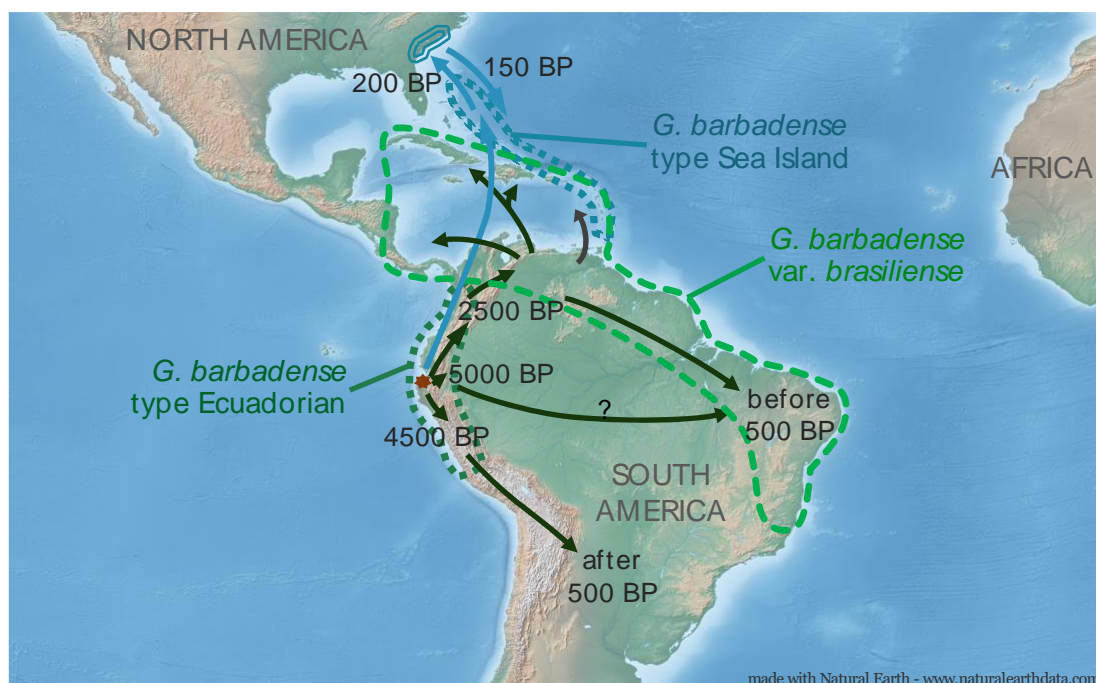


Figure 11. Hypothetical dispersals of *G. barbadense* from a domestication center in northwestern Peru-southwestern Ecuador and main geographic areas of the primitive domesticated forms. Sources: Hutchinson *et al.*, 1947; Stephens, 1973; Lazo, 1991; Percy, 2009; Ding *et al.*, 2015.

The timeline of the initial dispersals from the main center of diversity in northwestern Peru-southwestern Ecuador remains uncertain. It is unknown when *G. barbadense* was first cultivated in coastal valleys of Colombia, or was disseminated into the Amazonas, or into northern South America and the Caribbean. *Gossypium barbadense* cotton was certainly cultivated in coastal southern Peru by 4,500 YBP, in coastal Colombia by 2,500 YBP or earlier, and in northeastern Brazil before European colonization. *Gossypium barbadense* cotton was, from ca 5,000 to 3,000 YBP an important crop in the Marañón River valley along the eastern base of the Cordillera of the Andes (Bird, 1964; Lazo, 1991). Most probably the dispersal to Paraguay and Argentina was post-Columbian, as these cottons show traits of modern germplasm, but from West Andean origin and not from the extra-long *sea island* or modern improved forms (Percy and Wendel, 1990).

Three main formal or informal infraspecific types or varieties are recognized within *G. barbadense* (Table 3); *G. barbadense* var. *brasiliense* is the only formally recognized infraspecific entity. A very close sister-species of *G. barbadense*, *G. darwinii*, is endemic to the Galapagos Islands (see Figure 3) where it grows in truly wild populations. *G. darwinii* was considered as a local form or subspecies of *G. barbadense*, and even as a possible wild progenitor of the domesticated *G. barbadense* (Fryxell, 1979), until molecular genetics data showed that it is neatly distinct from *G. barbadense* (Wendel and Percy, 1990; Wu *et al.*, 2018; Chen *et al.*, 2020).

Table 3. *Gossypium barbadense* intraspecific diversification. Sources: Bauhin, 1623; Linnaeus, 1753; Hutchinson *et al.*, 1947; Hutchinson, 1951; Stephens, 1973; Percy and Wendel, 1990; Ding *et al.*, 2015; Castro *et al.*, 2016.

Race	Region	Description	Comments
Wild	Guayas region of northern Peru and southern Ecuador	Perennial shrub.	
<i>Gossypium barbadense</i> type Ecuadorian	Peru, Ecuador	Perennial shrub. Tolerant of extremely arid conditions. Seeds never fused together. Fiber more or less light grey, coarse with medium staple length. Cultivated varieties Tanguis and Aspero of Peru, Quebradinho of Brazil.	These are included within <i>Gossypium barbadense</i> var. <i>barbadense</i> .
<i>Gossypium barbadense</i> var. <i>brasiliense</i>	Amazon and Orinoco basins	Perennial shrub. Unique among cultivated cottons by its adaptation to <i>humid</i> tropical conditions. Morphologically variable but no clear-cut regional differentiations. Tendency of long narrow bolls; seeds fused together forming a mass, called «kidney seed», in regionally varying percent of the plants. Fiber fine and rather long (25-32mm).	The type of first botanical descriptions of South American cotton.
<i>Gossypium barbadense</i> type Sea Island	West Indies, Atlantic coasts of South Carolina, Georgia and Florida	Perennial or annual shrub. Selected c. 1780 from finer <i>G. barbadense</i> types introduced from western South America. Annual habit acquired when cultivated in coastal South Carolina.	Prominently involved in the modern cultivated <i>G. barbadense</i> germplasm of Egyptian and Pima cottons.

a. G. barbadense var. *brasiliense*

This botanical variety is distributed in northeastern South America and the Amazon Basin. It is particularly characterized by an adaptation to humid tropical conditions. It also shows a variable percent of "kidney seed" cotton plants, known as kidney-cotton, chain-cotton, or stone-cotton (Lewton, 1920), which inspired the name *G. lapideum* (Table 3), in which the seeds of each locule are aggregated into a kidney-shaped mass which permits a more easy hand-ginning than with independent seeds, by holding the block of seeds of one locule to separate the fiber (Stephens, 1967; Fryxell, 1979). No longer an advantageous trait, it persists only as a remnant. The kidney seed trait is monogenic and recessive (Turcotte and Percy, 1990). Globally, there appeared no sharp genetic distinction of *G. barbadense* var. *brasiliense* from other *G. barbadense* and it is now considered a locally adapted cultigen rather than a distinct species or subspecies.

b. G. barbadense type *Sea Island*

The Sea Island cottons were developed on the coastal islands - wherefrom the name of these cottons, while *G. hirsutum*, grown inland, was characterized Upland cotton - of Georgia and South Carolina (Smith and Cantrell, 1999). Historical data regarding the precise origin of the Sea Island cottons are limited. It was suggested that they originated from previous Sea Island cottons derived from *G. barbadense* var. *brasiliense* dispersals in the Caribbean (Stephens, 1976), while molecular genetic data (Percy and Wendel, 1990; Ding *et al.*, 2015) support the hypothesis of an original introduction from west Andean Peruvian stocks. Thus, while *G. barbadense* var. *brasiliense* was widely disseminated in the Caribbean, finer *G. barbadense* types from western South America were also introduced, probably no later than 1750 (Smith and Cantrell, 1999), and cultivated in northern West Indies where breeding made them evolve into the annualized, extra-long and fine Sea Island cotton. The intensive production of the exceptionally long and fine Sea Island cotton in southeastern United States on the Atlantic coasts of South Carolina, Georgia and Florida extended from 1785 to 1920. Present-day extra-long and extra-fine Sea Island varieties of the West Indies come from the germplasm of southeastern United States (Percy, 2009; Wendel *et al.*, 2010).

Genetic introgressions from *G. hirsutum* are considered to have been important in the development of Sea Island cottons, particular in enabling day-length neutrality and the characteristic extra-long fibers (Stephens, 1975; Percy, 2009). An experiment of hybridization of a *G. barbadense* genotype having a long growing cycle and short coarse fibers with a wild form of *G. hirsutum* having a long growing cycle and similarly short, but fine, fibers (Stephens, 1975) gave a progeny with fine and long fibers and a short growing cycle, and back-crosses successfully yielded a cotton with these favorable characteristics while morphologically appearing almost entirely *G. barbadense*. The hypothesis is that a similar event could have happened during the efforts of the cotton planters of southern Georgia, in the second half of the 18th century, to adapt *G. barbadense* to cultivation in these regions where perennial cottons couldn't be cultivated because of the cold winters. The emergence of the Sea Island cotton, including a history of the successive human-mediated dispersals from west-Andean Peruvian genotypes through to its distribution in the West Indies, has not yet been precisely established, but the essential points of its genetic origins now seem rather well-defined. It is possible that comparative whole genome sequencing data (Yuan *et al.*, 2021) will provide the clues necessary to tease out the genetic factors that were introgressed from *G. hirsutum* during its development.

c. Egyptian and Pima ELS Cottons

The modern elite, extra-long and extra-fine Egyptian and Pima cultivars that now dominate *G. barbadense* cultivation worldwide trace their origins to the Sea Island cottons (Percy and Wendel, 1990; Smith and Cantrell, 1999). The expansion of *G. barbadense* cultivation over the last two centuries was greatly facilitated by the genetic improvements contributed by Sea Island cultivars, which might thus be considered a pivotal step in the continued domestication of *G. barbadense*. Cotton varieties derived from the Ecuadorian type or var. *brasiliense* remain as door-yard cottons and still have an important role breeding programs, along with Sea Island germplasm (Castro *et al.*, 2016).

Modern extra-long staple (ELS) Egyptian cotton was created by selection from germplasm apparently stemming from natural cross-pollinations between a Sea Island lineage and a local perennial *G. barbadense* known as Jumel's cotton grown in Egypt since ca. 1817 (Watt, 1907; Percy and Wendel, 1990; Robert, 1997; Smith and Cantrell, 1999). In the southwestern USA, the extra-long staple cotton that became established at the beginning of the 20th century was primarily based on ELS germplasm introduced from Egypt. A schema capturing key components of this history leading to the globally dominant modern ELS cultivars is shown in Figure 12.

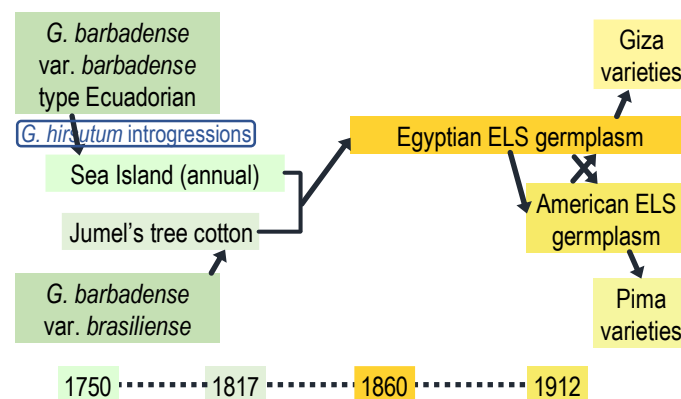


Figure 12. Schematic genealogy of the modern extra-long staple (ELS) Egyptian and American cotton gene pools. Sources: Watt 1907; Robert 1997; Smith and Cantrell 1999.

D - Domestication of *Gossypium hirsutum*

1. Genetic diversity in *G. hirsutum* and the concept of geographic or morphological races

With respect to *G. hirsutum*, it is first necessary to convey an appreciation for the incredibly wide geographic distribution of variously semi-wild, feral, and cultivated forms that existed at the time of European colonization. By this time, in addition to truly wild populations, various *G. hirsutum* plants spanning the full

wild-to-domesticated-to-feral continua were discovered to be broadly distributed from the SW United States through Central America, the Caribbean Islands, and even northern parts of S. America and adjacent islands, having disseminated to these regions over the millennia since initial domestication. Faced with this vast morphological and geographical diversity, it was unclear from the outset where initial domestication took place. Hutchinson (1951) after decades of comparative study and synthesis, published a seminal work wherein he defined seven “races” of *G. hirsutum*, one annualized form, race “*latifolium*”, that provided the foundation for the modern cotton crop that dominates global cotton commerce today, and six perennial races. We stress that these “races” are not taxonomic categories nor are they necessarily equivalent in terms of diversity or even biological reality. That is, the “races” may more properly be considered, using modern genetic understanding, as morphologically (not genetically) conceived multidimensional peaks of various heights and breadths in a continuous landscape or network of genetic diversity, as opposed to biologically distinct entities. Given the many centuries of crop improvement under directional selection, as well as many centuries of human-mediated germplasm transport and exchange by dozens of different cultures and peoples, it is unsurprising that clear cut categories of diversity within *G. hirsutum* simply do not exist. Yet Hutchinson’s “races” have some utility if understood in this modern multidimensional genetic sense.



Figure 13. **a.** *G. hirsutum* race *yucatanense* in flower, Guadeloupe. **b.** *Gossypium hirsutum* Upland field, United States. **c.** *Gossypium hirsutum* race *yucatanense* capsule and seed. Crane Point, Key Vaca, Florida, United States. **d.** *Gossypium hirsutum* race *yucatanense* vs cultivar Acala Maxxa seeds with fiber (lint lengths ca. 17 mm and 29 mm, respectively).

In agreement with classical taxonomic studies, molecular genetics establishes the wild race *yucatanense* as the ancestor of the domesticated germplasm and race *latifolium* as the direct ancestor of the Upland type of cotton now cultivated worldwide (see Fig. 13 for the most conspicuous transformations undergone by the wild *yucatanense* through to the modern cultivars). Relationships among the other “races”, however, remain hard to capture in simple trees or schematics, for the aforementioned reasons that the races are not really distinct nor uniform, but that they instead represent certain clusters of highly variable traits. This reality is captured in all modern molecular phylogenomic analyses (Wendel *et al.*, 1992; Brubaker and Wendel, 1994; Hinze *et al.*, 2016; Zhang *et al.*, 2019), most comprehensively and recently by Yuan *et al.*

(2021) who employed resequencing to study the entire genomes of 795 *G. hirsutum* and 201 *G. barbadense* accession spanning the full wild to domesticated continuum and including 247 and 114 wild or landrace accessions, respectively, in the two species. With respect to *G. hirsutum*, the data demonstrate a clear distinction between landrace forms from the Caribbean (Landrace 1) vs Central America/Mexico (Landrace 2), as well as the origin of the modern gene pool from within the Landrace 2. Importantly, though, phylogenetic and principal component analyses demonstrate complex and interdigitating patterns of diversity that aren't really congruent with neat separation or linear ordering into Hutchinson's geographic races. Nonetheless, for historical reasons and because the races retain a modicum of geo-morphological utility, we present a synopsis of these in Table 4, with updated comments that reflect modern phylogenomic analyses.

Table 4. Description of the *G. hirsutum* races. Sources: Hutchinson *et al.*, 1947; Hutchinson, 1951; Stephens, 1973; Brubaker and Wendel, 1994; Brubaker *et al.*, 1999; Wendel *et al.*, 2009; Zhang *et al.*, 2019; Yuan *et al.*, 2021.

Race	Varieties according to (Hutchinson <i>et al.</i> , 1947)	Races according to (Hutchinson, 1951)	Comments (Stephens, 1973; Brubaker <i>et al.</i> , 1999; Wendel <i>et al.</i> , 2009)
<i>yucatanense</i>		Wild. Perennial shrub. Northwestern coast of Yucatan peninsula.	Purely wild <i>G. hirsutum</i> only found in undisturbed beach strand vegetation. Highly-branched subshrub. Seed with hard, impermeable seed coat. Sparse, brown lint which is spinnable. Also found in Tamaulipas, Revillagigedo Island, Florida Keys, other dry areas in the Caribbean, etc.
<i>punctatum</i>	Cultivated (indigenes) in Tehuantepec Isthmus, Atlantic slope; also growing wild in Cuba, Haiti, Puerto Rico.	Perennial shrub. Yucatan peninsula and northward on Atlantic slope, to Florida (USA) and Bahamas.	Agronomically primitive domesticated race. Three-lobed leaves with shallow sinuses; bolls numerous but rather small and lint short. Wide human-mediated geographic dispersion.
<i>marie-galante</i>	Wild and cultivated in Caribbean region (Antilles), Ecuador, Guianas, northern and eastern Brazil (varietal type Moco)	Perennial. Tree form of growth. Northern Central America (Guatemala) southward to Colombia on both coasts, Caribbean region (Antilles) and northeastern Brazil.	Photoperiodic. Introgressed by <i>G. barbadense</i> according to molecular genetics data. 'Moco' cottons cultivated in the arid 'serido' region of northeastern Brazil on c. 500,000 ha until the 1980s.
<i>palmeri</i>		Perennial shrub. Pacific slope, southern Mexico west of Isthmus of Tehuantepec	Previously <i>G. lanceolatum</i> Tod. Pyramidally-shaped shrubs with lacinate leaves, prolific bloom and small bolls.
<i>morrilli</i>		Perennial shrub. Inland montane, southern Mexican plateau and northward.	Stout shrubs with numerous branches and small bolls.
<i>richmondi</i>		Perennial shrub. Pacific slope in Gulf of Tehuantepec region.	Large shrubs with numerous branches and medium-size bolls.
<i>latifolium</i>		Shrub, perennial or annual. Guatemala (both slopes) and southernmost Mexico (Chiapas), nearby areas. Hypothetically annualized with cultivation in plateau areas of Aztec state.	Short and compact morphology, large to vary large bolls, earliness, variable day-length sensitivity. Geographic range extending north to central Mexico and south to El Salvador and Nicaragua. Ancestor of modern Upland cultivated germplasm through its spread into present US cotton belt in early 17 th c.

The geographic distribution of these seven races is shown in Fig. 14, following (Hutchinson, 1951; Stephens, 1973). The wild race *yucatanense*, now represented by populations growing on well-drained and sandy soils on the northern shoreline of the Yucatan peninsula and other dry in the Caribbean, may represent the best-developed extant population system of truly wild *G. hirsutum* (Wendel *et al.*, 2009). Presumptively wild populations are found scattered on the drier, more undisturbed sites in many Caribbean Islands, the

Florida Keys, and northward along the Gulf Coast of southern Florida. The two widely cultivated primitive domesticated races *punctatum* and *marie-galante* now cover very wide areas. The other extant primitive domesticated *G. hirsutum* races *palmeri*, *morrilli* and *richmondi* show more or less restricted geographic extensions (Wendel *et al.*, 2009). Importantly, these geographical descriptions have undoubtedly been rendered out of date or even obsolete in the more than 70 years since Hutchinson formalized *G. hirsutum* diversity in this manner.

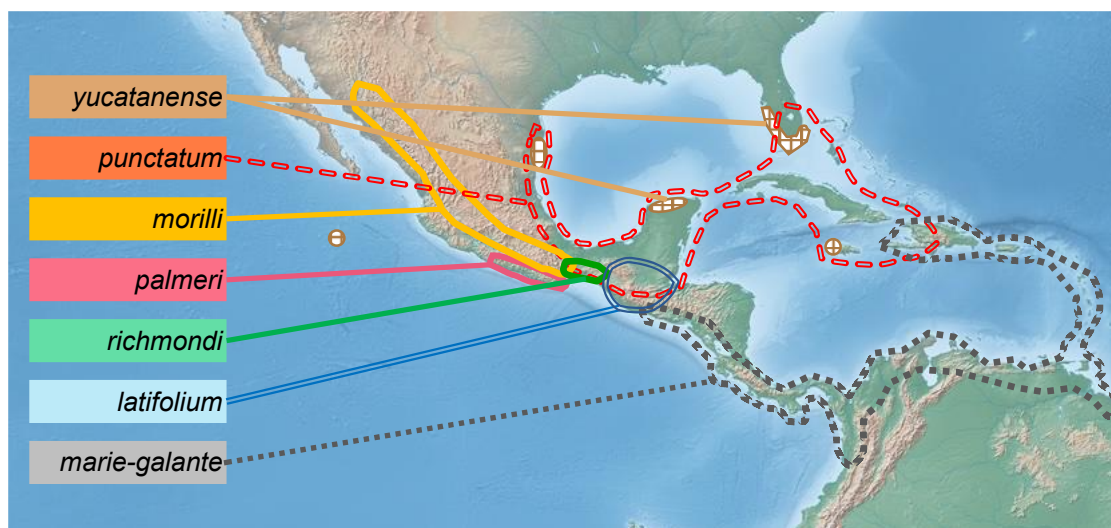


Figure 14. Hypothetical geographic distribution of the seven races of *G. hirsutum* at the beginning of European colonization. Sources: Hutchinson, 1951; Stephens, 1973.

The races *punctatum*, *marie-galante*, *palmeri*, *morrilli*, *richmondi* and *latifolium* show domestication traits simultaneously with some characteristics of wild cottons that are absent in modern cultivated *G. hirsutum*, including hard seed coats and delayed germination, short light cycle-linked fruiting, and greater disease resistance and drought tolerance. All wild forms of *G. hirsutum* produce a lint that is spinnable even while low in quality. Race *punctatum*, morphologically and agronomically the most primitive domesticated form of *G. hirsutum*, appeared as a possible ancestor of races *marie-galante*, *palmeri*, *morrilli*, *richmondi* and *latifolium* (Lubbers and Chee, 2009), but this simplistic linearity is not supported by modern phylogenomic data. Race *marie-galante* is a tree-like cotton and agronomically more distinct from the other *G. hirsutum* races than any other race. Molecular data demonstrated that *marie-galante* includes *G. barbadense* introgression, which could have occurred many times over hundreds or even thousands of years as the two species became sympatric, especially in the Caribbean. Race *latifolium*, the only day-length neutral annual among the primitive domesticates and the race morphologically and agronomically closest to modern Upland germplasm, appears as the direct ancestor of the latter (Stephens, 1973; Brubaker and Wendel, 1994; Lubbers and Chee, 2009; Yuan *et al.*, 2021).

In conclusion, the concept of *G. hirsutum* races has some, but limited modern utility. The ends of this spectrum of diversity, i.e., races *yucatanense* and *latifolium*, are perhaps the most useful in terms of defining gene pools, and race *marie-galante* continues to be distinctive, but the other perennial races are difficult to distinguish outside of their natural geographic context and seem unjustified with respect to patterns of diversity in a modern genetic sense.

2. Timeline and location of *G. hirsutum* domestication

The place of the initial domestication of *G. hirsutum* has been hypothesized as being either: 1) Southern Mexico and Guatemala (Hutchinson *et al.*, 1947), as a center of variability of *G. hirsutum*; 2) the Tehuantepec region (Stephens, 1973) on the basis of present-day distribution of the natural populations of the primitive domesticated races; 3) northeastern Brazil (Ano *et al.*, 1982), hypothetically also the center of origin of the whole Upland cultivar group; 4) the Yucatan peninsula (Wendel *et al.*, 1992, 2010; Brubaker and Wendel, 1994), the home of wild race *yucatanense* and where plants of race *punctatum*, considered the most primitive cultivated *G. hirsutum* race, are frequently encountered.

With respect to the question of the location of initial domestication, one criterion that might be useful is information provided by archaeological data. In Mesoamerica, the oldest cotton fibers associated with human settlements are dated to about 5,500 YBP, with cotton cultivation and woven cotton textiles appearing just before 3,000 YBP in various places over a vast region from central Mexico to the Maya area (part A of this chapter and Figure 8), thus providing little insight into particular regions for the initial domestication and cultivation of *G. hirsutum*. The situation has improved in recent decades with the application of modern genetic tools. Studies of variability in *G. hirsutum* identified two geographically broad centers of diversity, one in southern Mexico and Guatemala, and the other in the Caribbean (Wendel *et al.*, 1992). Although a center of genetic diversity need not correspond to the center of domestication, (Harlan, 1971), these data suggested that one or both of southern Mexico-Guatemala and the Caribbean could be implicated in the origins of *G. hirsutum* domestication.

Another potentially useful criterion in inferring geographic origins of domestication is the location of primitively domesticated forms. Two of Hutchinson's races cover wide geographic areas, race *punctatum* in the Yucatan Peninsula and in drier localities in parts of the Gulf of Mexico, and race *marie-galante* in Central America, the West Indies and the Atlantic coasts of northern South America (see Figure 14). Both of these races were widely disseminated as textile crops during the past, perhaps thousands of years; because they are so widely distributed geographically, however, they offer no insight into sites of initial domestication. The other primitive domesticated *G. hirsutum* races, i.e., *morilli*, *palmeri*, *richmondi* and *latifolium* occupied, at the time of their initial description (Hutchinson, 1951), more or less small areas along the Pacific side of Mexico to Guatemala, which could implicate this region as a site of initial domestication. The Pacific coast of Mexico and surrounding regions have a high environmental suitability for wild *G. hirsutum*, but so do other areas, for example, in the Yucatan and elsewhere in the Caribbean (Alavez *et al.*, 2021).

One widely cited hypothesis (Brubaker and Wendel, 1994; Lubbers and Chee, 2009; Yuan *et al.*, 2021) is that the northern Yucatan Peninsula was the site of *G. hirsutum* domestication, fitting with the observation of the wild race *yucatanense* growing on its northern shoreline and of plants of *punctatum*, the cultivated race considered most primitive, disseminated over the peninsula. Because no traits characteristic of domesticated cottons are found in race *yucatanense* (Stephens, 1973), and because of its location at the northern tip of a karstic region without standard access to drinking water, this region has also been considered to be unlikely as the site of initial domestication. A possibility is that cotton taken from a wild *yucatanense* population was brought to a more or less distant human settlement where the seeds separated from their fibers could have initiated cultivation and the domestication process independently from the wild population. The Yucatan Peninsula is a rather extended region - ca. 300 km wide by 600 km long - where two other locations show environmental suitability for wild *G. hirsutum* (Alavez *et al.*, 2021): 1) northern Belize lowlands where Maya agricultural settlements have been identified (Miksicek, 1990) and, 2) the Gulf Coast of eastern Tabasco close to Olmec (La Venta) and Maya (Aguada Fénix) archaeological sites (Pool, 2007; Inomata *et al.*, 2020).

Alternative proposals to the Yucatan Peninsula for *G. hirsutum* initial domestication include the Pacific coast from southern Mexico to Guatemala as it is favorable to *G. hirsutum* natural populations (Smith and Stephens, 1971; Alavez *et al.*, 2021). There, some wild *G. hirsutum* race *yucatanense* population established in the coastal regions could have been domesticated, but this also requires a subsequent extinction of the *yucatanense* population concomitant with or subsequent to domestication into primitive cultivated forms. On the coastal dunes of the Mexican Pacific coast that extends from the Tehuantepec isthmus are found *G. hirsutum* populations that are part of the natural flora, although these apparently are not the "truly wild" (Wegier *et al.*, 2011; Coppens d'Eeckenbrugge and Lacape, 2014; Alavez *et al.*, 2021; Wendel, personal observations) forms that would seem able to support this hypothesis.

A separate domestication that gave race *marie-galante* has been proposed (Stephens, 1973; Lubbers and Chee, 2009) in order to explain the strong dominance in the Caribbean of this race and its very distinctive morphological and phenological traits. Molecular genetics data, though, refute this hypothesis of independent domestication of race *marie-galante* (Wendel *et al.*, 1992; Brubaker and Wendel, 1994; Zhang *et al.*, 2019; Yuan *et al.*, 2021), which instead clearly has an introgressive ancestry with domesticated *G. barbadense* as a contributing partner. Thus, there remains no evidence of more than a single domestication of *G. hirsutum*.

In conclusion, much uncertainty remains about the precise initial place of *G. hirsutum* domestication, although genetic data combined with present-day distribution of wild forms support the Yucatan as the most probable location (Yuan *et al.*, 2021). Perhaps future archaeological studies will shed more light on this problem. The oldest cotton fibers in Mesoamerican archaeological sites are dated to *circa* 5,500 YBP, so perhaps cotton domestication began at about or not much before 5,500 YBP. Cotton fabrics, small spinning whorls adequate for cotton and hints of cotton cultivation appear only in the following centuries in adjacent regions, mainly between 3,000 YBP and 2,000 YBP, such that the first suitable cotton varieties were probably created not much time before 3,000 YBP. It is, for instance, in the Middle Formative, 3,000 to 2,400 YBP that cotton textiles are evidenced in the Olmec heartland (Follensbee, 2008). This archaeological record is informative, but of course incomplete, so we await new insights from the earliest period of cotton domestication.

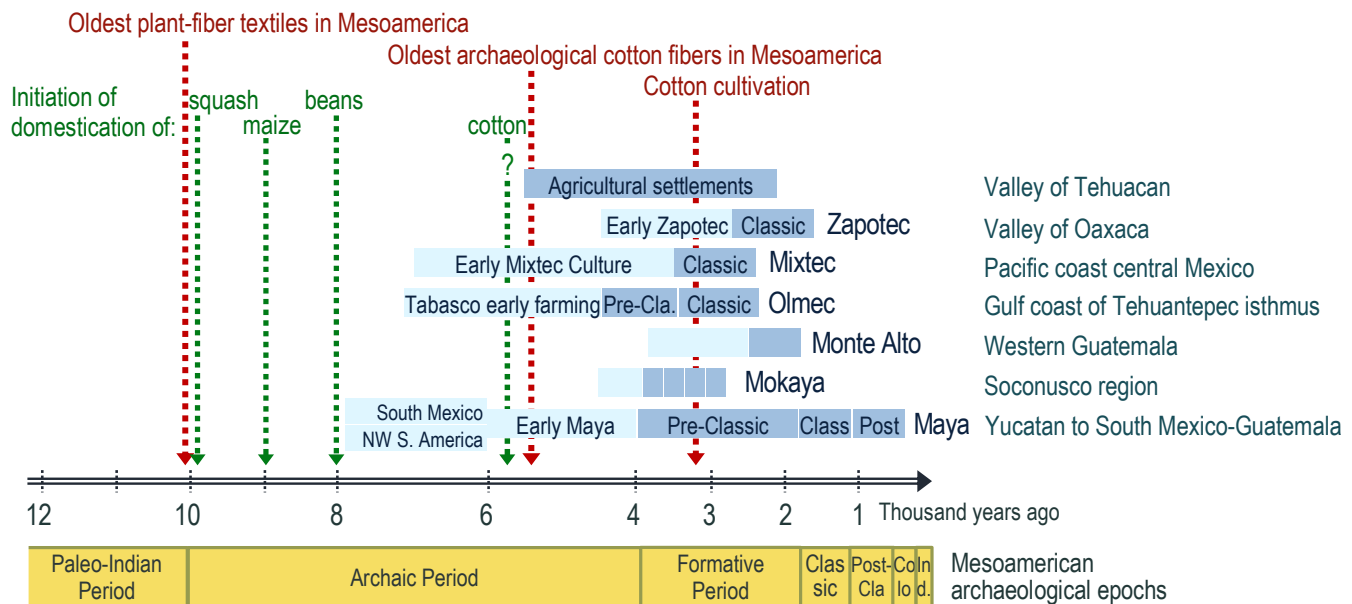


Figure 15. Chronology of earliest Mesoamerican crop domestications, main civilizations and first cotton textiles and cotton cultivation, including an hypothetical time for *G. hirsutum* domestication. Pre-Cla=Pre-Classic, Class=Classic, Post=Post-Classic, Colon.=Colonial era, Ind.=Independencies. Sources: MacNeish, 1967a; Smith, 1997; Arnaiz-Villena *et al.*, 2004; Lohse *et al.*, 2006; Pool, 2007; Follensbee, 2008; Estrada-Belli, 2010; Zizumbo-Villarreal and Colunga-GarcíaMarín, 2010; Rendón-Anaya *et al.*, 2017; Stitzer and Ross-Ibarra, 2018; Kennett *et al.*, 2020, 2022.

Based on our present knowledge on the timeline of earliest crop domestications and of civilization development in Mesoamerica (Figure 15), the beginning of *G. hirsutum* domestication, hypothetically between 6,000 and 5,500 YBP, would have followed by two to four millennia those of maize, beans and squash, but it also would predate by one to two millennia the Formative periods of the main civilizations, Olmec and Maya. No developed civilization seems to deserve credit for the domestications of squash, maize and beans, and for cotton it could correspond to early Mixtec culture in southwestern Mexico, to early Olmec culture in southwestern Yucatan Peninsula or early Maya culture in the region from central Guatemala to Belize. For this period in the Archaic times of Mesoamerica, the description proposed by (MacNeish, 1967a) is of small human groups of a few families. A recent study (Kennett *et al.*, 2022) proposes that the hypothetical time of initial cotton domestication follows a human dispersal into southeastern Yucatan, recently dated between 7,300 and 5,600 YBP by genome-wide ancient DNA analyses. The arrival of those immigrant people genetically related to present-day populations living in regions extending from Costa Rica to Colombia was followed by a period of new horticultural practices and marked development of agriculture in the Maya region from *ca.* 5,000 YBP, according to these authors. The estimated time period when these migrants arrived in southeastern Yucatan is thus *ca.* 1,500 to 0 years prior to the possible origin of *G. hirsutum* domestication. We raise the possibility that the immigrant people brought from northwestern South America, along with better maize varieties, some knowledge related to the textile use of *G. barbadense* cotton and initiated the textile use and domestication of *G. hirsutum*. Crop exchanges and dispersals between

South America and Mesoamerica occurred early along the millennia of prehistoric agriculture: corn, domesticated in Mesoamerica, was cultivated prior to 7,500 YBP in South America (Pagán-Jiménez *et al.*, 2015), while manioc and cacao, domesticated in northern South America, were cultivated in Mesoamerica at ca. 6,500 YBP (Isendahl, 2011) and ca. 4000 YBP (Zarrillo *et al.*, 2018), respectively. Cotton cultivation and use was well developed in northern South America at the time when migrants from there came to Mesoamerica, as explained above, and their knowledge could have been applied to the wild *G. hirsutum* cotton encountered in the southern Mexico-Guatemala region or the Yucatan Peninsula.

The history of the domestication of the diploid African-Asian cotton species *G. arboreum* and *G. herbaceum* shows that between the oldest archaeological remains with cotton fibers - logically harvested from wild, instead of cultivated, plants - and the spread of cultivation of cotton and elaboration of cotton textiles, timespans up to thousands of years have passed (Viot, 2019), probably because the wild cotton plant still could not be easily cultivated before some genetic adaptation. We note that this is likely a slow process in *Gossypium*, because of its relatively long life-cycle and woody perennial nature. For plants with these traits, the fitness gains (in a crop domestication sense) from strong directional human-mediated selection are expected to be much slower than for annual plants such as maize, squash and beans. Also, in comparison to the leaf or bast fibers from *Agavaceae*, *Bromeliaceae* and *Cyperaceae* that were used before cotton (Jolie *et al.*, 2011), the cotton fiber is short and its textile use implies its spinning, which must have been a technical challenge mastered only by people already competent at spinning and weaving other textile fibers (Hutchinson *et al.*, 1947; Beresford-Jones *et al.*, 2018); making yarns by spinning appeared rather late along the timeline of textile craftsmanship in Mesoamerica, and probably was there an interaction with the textile use of *G. hirsutum* cotton.

E – Post-domestication Interspecific Introgression between *Gossypium hirsutum* and *G. barbadense*

Following their parallel initial domestication in different continents, *G. hirsutum* and *G. barbadense* began to become sympatric in parts of their increasingly broad cultivated ranges, thanks to human-mediated dispersal of primitive domesticated types, over long periods of time in many regions in Central America, the Caribbean and northern and northeastern South America. This sympatry created frequent opportunities for hybridization between the two species. It has been postulated that interspecific introgressions from *G. hirsutum* were essential for the improvement of *G. barbadense* into the daylength-neutral, annual, extra-long fiber types that gave rise to the modern Sea Island types and ultimately the Egyptian and Pima cottons (Stephens, 1976). Molecular studies have confirmed this introgression (Brubaker *et al.*, 1993; Fang *et al.*, 2017; Hu *et al.*, 2019; Yuan *et al.*, 2021). In return, *G. barbadense* introgression into *G. hirsutum* has been considered important in race *marie-galante* (Stephens, 1967, 1973; Wendel *et al.*, 1992).

Despite the sympatry and the interspecific hybridizations resulting in these recorded introgressions, *G. hirsutum* and *G. barbadense* are maintained separately as distinct crop species, an independence which genetic barriers inducing a partial reproductive isolation can explain. A phenomenon that has been known for nearly a century or perhaps longer is that of “F2 breakdown” (Harland, 1936; Stephens, 1949), the observation that interspecific *G. hirsutum* x *G. barbadense* F1 hybrids are readily obtained and fully fertile, but that selfed generations from the F2 on show high frequencies of defective recombinant plants, whose elimination progressively leads to the recovery of parental types. No full sterility thus exists, but few interspecific recombinants only are viable, limiting gene flows and maintaining the specificity of *G. hirsutum* and *G. barbadense*. Chromosomal inversions totaling 170.2 Mb of genome sequence over 15 of the 26 chromosomes differentiate *G. hirsutum* and *G. barbadense* (Wang *et al.*, 2019). As a comparison, the total length of chromosomal inversions between two rather distant *G. hirsutum* cultivars, the American TM-1 and the Chinese ZM24, is 43.2 Mb (Yang *et al.*, 2019). Chromosomal inversions are common in plants and are considered to play a role in speciation (Huang and Rieseberg, 2020), and also subsequent isolation.

From a modern molecular genetic perspective, there are many possibilities for F2 or later generation “breakdown” or incompatibility, including the disruption of coadapted gene complexes by recombination (Jiang *et al.*, 2000; Chen *et al.*, 2020), divergent neo- and sub-functionalization of homoeologous genes (Viot *et al.*, 2009; Hu *et al.*, 2019; Birchler and Yang, 2022) disruptions of gene balance (Birchler and Veitia, 2021; Birchler and Yang, 2022), or many other forms of regulatory mismatch (Hu *et al.*, 2019). Irrespective of the actual causes of regulatory or developmental deficiency, it is interesting that the two species are still able to easily hybridize after an estimated 500,000 years of divergence and independent evolution (Pan *et al.*, 2020).

Of note, although partial reproductive isolation would appear to limit interspecific gene exchanges, over time, these reciprocal introgressions have been extensive, as convincingly demonstrated by genome sequencing data (Yuan *et al.*, 2021).

The reproductive barrier between *G. hirsutum* and *G. barbadense* might have evolved for the most part as a consequence of the accumulation in isolation, over time, of many genetic and genomic differences, instead of as a result of mechanisms specifically preventing interspecific hybridizations and gene flow during recent sympatry. Intriguingly, the relatively recent sympatry following domestication has led to the evolution of a likely example of the latter, “reinforcing” form of species isolation (Stephens and Phillips, 1972). In this system, certain alleles at the “corky” locus are only found in areas of sympatry; when combined, unfit, abnormal corky phenotypes are produced, thus reinforcing species isolation. The latter agrees with the hypothesis of genes, not large chromosomal rearrangements, as typical systems that are selected for postzygotic isolation (Coyne and Orr, 1998). Thus, *G. barbadense* and *G. hirsutum* provide a fascinating example of a wealth of mechanisms involved in plant speciation and diversification.

IV - Conclusions

Here we have provided a synthesis of our knowledge regarding *Gossypium* evolution and domestication of the American allopolyploid species, based on decades of research and integrating data from fundamental taxonomic investigations, biogeography, molecular genetics, phylogenetic analysis and archaeology. These diverse and extensive sources of information reveal a rich depiction of the natural and human history of the genus, provide a temporal perspective on diversification and polyploid formation, uncover cryptic interspecific hybridizations, clarify and contribute to the taxonomy of the genus, and offer a firm foundation for understanding parallel cotton domestications in Mesoamerica and South America. This information is vitally important to our understanding of the genomic architecture of the world's most important fiber plant, and contributes substantially to our understanding of general biological principles.

With respect to evolutionary phenomena, a truly remarkable feature of the genus is the high frequency of both long-distance dispersal and interspecific hybridization, the latter often between lineages that one would surmise have little opportunity for hybridization. This important aspect of *Gossypium* evolution, noted also by earlier authors (Cronn and Wendel, 2004; Wendel and Grover, 2015; Hu *et al.*, 2021), is in evidence throughout the genus, most famously by the intercontinental, intergenomic hybridization that gave rise to the American allopolyploid species, which combine two highly divergent genomes (A and D), but also in the multiple examples of gene flow reviewed here, in both wild species, e.g., *G. gossypoides* and *G. aridum*, and with respect to the development of the important modern cultivated species *G. hirsutum* and *G. barbadense*. Thus, *Gossypium* offers a fascinating testimonial example of the importance and utility of fundamental botanical discovery combined with modern technological prowess to generate genomic insights. It also illustrates the reality that evolutionary history often may be more accurately represented by phylogenetic networks than by trees (Moret *et al.*, 2004; Bastide *et al.*, 2018; Wen *et al.*, 2018).

We also review the current state of our knowledge regarding the history of cotton domestication and diffusion in the Americas, an additional remarkable and seemingly unlikely story entailing parallel origins and parallel directional selection, transforming two different wild short-day perennial shrubs having small capsules and seeds covered by short, tan-colored epidermal trichomes into modern daylength-neutral annuals bearing abundant, fine, strong white fibers. This dual domestication was followed much later by unintentional and more recently intentional interspecific introgression, as the two species came into contact following their initial domestication in different hemispheres, and continued into the modern era following European colonization of the Americas. We review the archaeological literature bearing on the textile use of cotton lint, showing that it probably began as early as 8,000 years ago for the South American species *G. barbadense*, and almost incredibly, half a world away from a similar domestication process for the Asiatic diploid cotton *G. arboreum* in the Indus Valley of southern Asia. Textile use of cotton of the species *G. hirsutum* in Mesoamerica seems to have begun later, possibly ca. 5,500 YBP and could have taken ca. two millennia before the development of cotton cultivation. We suggest the hypothesis that human migration from northern South America to Mesoamerica might have a link to cotton cultivation. This intriguing possibility, i.e., that the domestication of *G. barbadense* provided the impetus for a much-later *G. hirsutum* domestication following human migration and cultural diffusion awaits additional evaluation from future archaeological discoveries.

Acknowledgments

The authors wish to thank David Pot, Head of the team GIV, UMR Agap, Cirad, for valuable comments. Jonathan Wendel's research over the past 35 years has been generously supported by funding agencies including the National Science Foundation of the USA and Cotton Incorporated.

Declarations

No conflicts of interest declared.

Copyrights

Photos: **8.a.** © JFW; **8.b.** © Forest and Kim Starr. License: CCA 3.0 us, Source: [https://commons.wikimedia.org/wiki/File:Starr-141217-5564-Gossypium_tomentosum-habit-Kaukaupapa-Kahoolawe_\(25223173406\).jpg](https://commons.wikimedia.org/wiki/File:Starr-141217-5564-Gossypium_tomentosum-habit-Kaukaupapa-Kahoolawe_(25223173406).jpg); **8.c.** © Salompillajo, License: CC BY-NC, Source: <http://www.inaturalist.org/observations/21021915>. **8.d.** © Bärbel Miemietz, License: CC BY-SA 4.0, Source: <https://commons.wikimedia.org/w/index.php?curid=119307714>; **8.e.** © Joseph P. Gallagher; **13.a.** © Georges Ano; **12.b.** © Jim Black, Free use, Source: <https://pixabay.com/photos/cotton-cotton-field-white-farming-4649804/> **12.c, d.** © JFW.

References

- Alavez, V., Cuervo-Robayo, Á.P., Martínez-Meyer, E., and Wegier, A. 2021. Eco-geography of feral cotton: a missing piece in the puzzle of gene flow dynamics among members of *Gossypium hirsutum* primary gene pool. *Front. Ecol. Evol.* **9**: 653271. doi: 10.3389/fevo.2021.653271
- Alvarez, I., and Wendel, J.F. 2006. Cryptic interspecific introgression and genetic differentiation within *Gossypium aridum* (Malvaceae) and its relatives. *Evol. Int. J. Org. Evol.* **60**: 505–517. <https://www.jstor.org/stable/4095313>
- Ano, G., Schwendiman, J., Fersing, J., and Lacape, M. 1982. Les cotonniers primitifs *G. hirsutum* race *yucatanense* de la Pointe des Chateaux en Guadeloupe et l'origine possible des cotonniers tetraploides du Nouveau Monde. *Coton Fibres Trop.* **37**: 327–332.
- Arnaiz-Villena, A., Gomez-Casado, E., Moscoso, J., Zamora, J., Martin-Villa, J.M., Perez-Blas, M., Lopez-Santalla, M., Lucas-Gramajo, P., Silvera, C., Lowy, E., *et al.* 2004. Origin of Mayans according to HLA genes and the uniqueness of Amerindians. *Hum. Immunol.* **65**: S112. doi: 10.1016/j.humimm.2004.07.213
- Barnard-Kubow, K.B., So, N., and Galloway, L.F. 2016. Cytonuclear incompatibility contributes to the early stages of speciation. *Evolution* **70**: 2752–2766. doi: 10.1111/evo.13075
- Bastide, P., Solís-Lemus, C., Kriebel, R., William Sparks, K., & Ané, C. 2018. Phylogenetic comparative methods on phylogenetic networks with reticulations. *Syst. biol.* **67**(5): 800–820. doi: 10.1093/sysbio/syy033
- Bauhin, C. 1623. *Pinax theatri botanici*. Basel.
- Beasley, J. O. 1940. The origin of American tetraploid *Gossypium* species. *Am. Nat.* **74**: 285–286.
- Beckert, S. 2014. *Empire of cotton: a global history*. Vintage, New York, USA.
- Bellaloui, N., Stetina, S.R., and Turley, R.B. 2015. Cottonseed protein, oil, and mineral status in near-isogenic *Gossypium hirsutum* cotton lines expressing fuzzy/linted and fuzzless/linted seed phenotypes under field conditions. *Front. Plant Sci.* **6**: 137. doi: 10.3389/fpls.2015.00137
- Bennetzen, J.L., and Kellogg, E.A. 1997. Do plants have a one-way ticket to genomic obesity? *Plant Cell* **9**(9): 1509–1514. doi: 10.1105/tpc.9.9.1509
- Beresford-Jones, D., Pullen, A., Chauca, G., Cadwallader, L., García, M., Salvatierra, I., Whaley, O., Vásquez, V., Arce, S., Lane, K., *et al.* 2018. Refining the maritime foundations of Andean civilization: how plant fiber technology drove social complexity during the Preceramic period. *J. Archaeol. Method Theory* **25**: 393–425. doi: 10.1007/s10816-017-9341-3
- Birchler, J.A., and Veitia, R.A. 2021. One hundred years of gene balance: how stoichiometric issues affect

- gene expression, genome evolution, and quantitative traits. *Cytogenet. Genome Res.* **161**: 529–550. doi: 10.1159/000519592
- Birchler, J.A., and Yang, H. 2022. The multiple fates of gene duplications: deletion, hypofunctionalization, subfunctionalization, neofunctionalization, dosage balance constraints, and neutral variation. *Plant Cell* **34**: 2466–2474. doi: 10.1093/plcell/koac076
- Bird, J.B. 1948. Preceramic Cultures in Chicama and Virú. *Memoirs of the Society for American Archaeology*, **4**: 21–28. doi:10.1017/S0081130000000320
- Brown, M.S. 1951. The spontaneous occurrence of amphiploidy in species hybrids of *Gossypium*. *Evolution* **5**: 25. doi: 10.2307/2405428
- Brown, M.S., and Menzel, M.Y. 1952. Polygenomic hybrids in *Gossypium*. I. Cytology of hexaploids, pentaploids and hexaploid combinations. *Genetics* **37**: 242–263. doi: 10.1093/genetics/37.3.242
- Brubaker, C.L., Bourland, F.M., and Wendel, J.F. 1999. The origin and domestication of cotton. In: Smith, C.W., and J.T. Cothren, Eds., *Cotton: origin, history, technology and production*. New York: John Wiley & Sons, pp. 23–32.
- Brubaker, C.L., Koontz, J.A., and Wendel, J.F. 1993. Bidirectional cytoplasmic and nuclear introgression in the New World cottons *Gossypium barbadense* and *G. Hirsutum* (Malvaceae). *Am. J. Bot.* **80**: 1203–1208. doi: 10.1002/j.1537-2197.1993.tb15353.x
- Brubaker, C.L., and Wendel, J.F. 1994. Reevaluating the origin of domesticated cotton *Gossypium hirsutum* (Malvaceae) using nuclear restriction fragment length polymorphisms (RFLP). *Am. J. Bot.* **81**: 1309–1326. doi: 10.1002/j.1537-2197.1994.tb11453.x
- Campbell, L. 1997. *American Indian languages: the historical linguistics of Native America*. New York: Oxford University Press. doi: 10.1017/S0047404500321030
- Carvalho, M.R., Herrera, F.A., Jaramillo, C.A., Wing, S.L., and Callejas, R. 2011. Paleocene Malvaceae from northern South America and their biogeographical implications. *Am. J. Bot.* **98**: 1337–1355. doi: 10.3732/ajb.1000539
- Castro, A.A., Hoffmann, L.V., Lima, T.H., Oliveira, A.I.D., Brito, R.R., Mendes, L. de M.O., Pereira, C.C.O., Malafaia, G., and Menezes, I.P.P. de 2016. *Gossypium barbadense*: an approach for in situ conservation in Cerrado, Brazil. *J. Agric. Sci.* **8**: 59. doi: 10.5539/jas.v8n8p59
- Chen, Z., Feng, K., Grover, C.E., Li, P., Liu, F., Wang, Y., Xu, Q., Shang, M., Zhou, Z., Cai, X., et al. 2016. Chloroplast DNA structural variation, phylogeny, and age of divergence among diploid cotton species. *PLoS ONE* **11**: e0157183. doi: 10.1371/journal.pone.0157183
- Chen, Z.J., Sreedasyam, A., Ando, A., Song, Q., De Santiago, L.M., Hulse-Kemp, A.M., Ding, M., Ye, W., Kirkbride, R.C., Jenkins, J., et al. 2020. Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement. *Nat. Genet.* **52**: 525–533. doi: 10.1038/s41588-020-0614-5
- Chen Z., Zhao N., Li S., Grover C. E., Nie H., Wendel J. F., and Hua J. 2017. Plant mitochondrial genome evolution and cytoplasmic male sterility. *Crit. Rev. Plant Sci.*, **36**(1): 55-69. doi: 10.1080/07352689.2017.1327762
- Coppens d'Eeckenbrugge, G., and Lacape, J.-M. 2014. Distribution and differentiation of wild, feral, and cultivated populations of perennial Upland cotton (*Gossypium hirsutum* L.) in Mesoamerica and the Caribbean. *PLoS ONE* **9**: e107458. doi: 10.1371/journal.pone.0107458
- Coyne, J.A., and Orr, H.A. 1998. The evolutionary genetics of speciation. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **353**: 287–305. doi: 10.1098/rstb.1998.0210
- Cronn, R., Small, R.L., Haselkorn, T., and Wendel, J.F. 2003. Cryptic repeated genomic recombination during speciation in *Gossypium gossypoides*. *Evolution* **57**: 2475–2489. doi: 10.1111/j.0014-3820.2003.tb01493.x
- Cronn, R., and Wendel, J.F. 2004. Cryptic trysts, genomic mergers, and plant speciation. *New Phytol.* **161**: 133–142. doi: 10.1111/j.1469-8137.2004.00947.x
- Cronn, R.C., Small, R.L., Haselkorn, T., and Wendel, J.F. 2002. Rapid diversification of the cotton genus (*Gossypium* : Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. *Am. J. Bot.* **89**: 707–725. doi: 10.3732/ajb.89.4.707
- Damp, J.E., and Pearsall, D.M. 1994. Early cotton from coastal Ecuador. *Econ. Bot.* **48**: 163–165. doi: 10.1007/BF02908209
- DeJoode, D.R., and Wendel, J.F. 1992. Genetic diversity and origin of the Hawaiian Islands Cotton,

- Gossypium tomentosum*. *Am. J. Bot.* **79**: 1311–1319. doi: 10.1002/j.1537-2197.1992.tb13735.x
- Denham, H.J. 1924. The cytology of the cotton plant I. Microspore formation in Sea Island Cotton. *Ann. Bot.* **os-38**: 407–432. doi: 10.1093/oxfordjournals.aob.a089908
- Dillehay, T.D., Bonavia, D., Goodbred, S.L., Pino, M., Vásquez, V., and Tham, T.R. 2012. A late Pleistocene human presence at Huaca Prieta, Peru, and early Pacific coastal adaptations. *Quat. Res.* **77**: 418–423. doi: 10.1016/j.yqres.2012.02.003
- Dillehay, T.D., Ramírez, C., Pino, M., Collins, M.B., Rossen, J., and Pino-Navarro, J.D. 2008. Monte Verde: seaweed, food, medicine, and the peopling of South America. *Science* **320**: 784–786. doi: 10.1126/science.1156533
- Dillehay, T.D., Rossen, J., Andres, T.C., and Williams, D.E. 2007. Preceramic adoption of peanut, squash, and cotton in northern Peru. *Science* **316**: 1890–1893. doi: 10.1126/science.1141395
- Ding, M., Ye, W., Lin, L., He, S., Du, X., Chen, A., Cao, Y., Qin, Y., Yang, F., Jiang, Y., et al. 2015. The hairless stem phenotype of cotton (*Gossypium barbadense*) is linked to a *Copia*-like retrotransposon insertion in a *Homeodomain-Leucine Zipper* gene (*HD1*). *Genetics* **201**: 143–154. doi: 10.1534/genetics.115.178236
- Endrizzi, J.E., Turcotte, E.L., and Kohel, R.J. 1985. Genetics, cytology, and evolution of *Gossypium*. *Advances in genetics* **23**: 271–375. doi: 10.1016/S0065-2660(08)60515-5
- Estrada-Belli, F. 2010. *The first Maya civilization. Ritual and Power Before the Classic Period*. Abingdon & New York: Routledge. doi: 10.4324/9780203839133
- Fang, L., Gong, H., Hu, Y., Liu, C., Zhou, B., Huang, T., Wang, Y., Chen, S., Fang, D.D., Du, X., et al. 2017. Genomic insights into divergence and dual domestication of cultivated allotetraploid cottons. *Genome Biol.* **18**: 33. doi: 10.1186/s13059-017-1167-5
- FAO 2021. Recent trends and prospects in the world cotton market and policy developments. Food and Agriculture Organization of the United Nations, Markets and Trade Division - Economic and Social Development Stream Markets. <https://www.fao.org/3/cb4589en/cb4589en.pdf>
- Feng, C., Ulloa, M., Perez-M., C., and Stewart, J.McD. 2011. Distribution and molecular diversity of arborescent *Gossypium* species. *Botany* **89**: 615–624. doi: 10.1139/b11-042
- Fillooy Nadal, L. 2017. Mesoamerican archaeological textiles: an overview of materials, techniques, and contexts. . In: Bjerregaard Lena and Peters Ann, *PreColumbian Textile Conference VII / Jornadas de Textiles PreColombinos VII*. Zea Books, Lincoln, NE. Pp. 7-39. *Zea Books*. doi: 10.13014/K2RB72S4
- Follensbee, B.J.A. 2008. fiber technology and weaving in Formative-Period Gulf Coast cultures. *Anc. Mesoam.* **19**: 87–110. doi: 10.1017/S0956536108000229
- Francis, D., Fonseca, R., Nelli, N., Cuesta, J., Weston, M., Evan, A., and Temimi, M. 2020. The atmospheric drivers of the major Saharan dust storm in June 2020. *Geophysical Research Letters* **47**: e2020GL090102. <https://doi.org/10.1029/2020GL090102>
- Fryxell, P.A. 1968. A Redefinition of the tribe Gossypieae. *Bot. Gaz.* **129**: 296–308. doi: 10.1086/336448
- Fryxell, P.A. 1971. Phenetic analysis and the phylogeny of the diploid species of *Gossypium* L. (Malvaceae). *Evolution* **25**: 554–562. doi: 10.1111/j.1558-5646.1971.tb01916.x
- Fryxell, P.A. 1976. *A nomenclator of Gossypium: the botanical names of cotton*. Agricultural Research Service, US Department of Agriculture, Hyattsville, Md.,USA.
- Fryxell, P.A. 1979. *The natural history of the cotton tribe (Malvaceae, tribe Gossypieae)*. College Station: Texas A&M University Press.
- Fryxell, P.A. 1992. A revised taxonomic interpretation of *Gossypium* L. (Malvaceae). *Rheede* **2**: 108–165.
- Gallagher, J.P., Grover, C.E., Rex, K., Moran, M., and Wendel, J.F. 2017. A new species of cotton from Wake Atoll, *Gossypium stephensii* (Malvaceae). *Syst. Bot.* **42**: 115–123. doi: 10.1600/036364417X694593
- Gao, W., Xu, F., Long, L., Li, Y., Zhang, J., Chong, L., Botella, J.R., and Song, C. 2020. The gland localized CGP1 controls gland pigmentation and gossypol accumulation in cotton. *Plant Biotechnol. J.* **18**: 1573–1584. doi: 10.1111/pbi.13323
- Gross, B.L., and Strasburg, J.L. 2010. Cotton domestication: dramatic changes in a single cell. *BMC Biol.* **8**: 137. doi: 10.1186/1741-7007-8-137
- Grover, C.E., Arick, M.A., Conover, J.L., Thrash, A., Hu, G., Sanders, W.S., Hsu, C.-Y., Naqvi, R.Z., Farooq, M., Li, X., et al. 2017. Comparative genomics of an unusual biogeographic disjunction in the

- cotton tribe (Gossypieae) yields insights into genome downsizing. *Genome Biol. Evol.* **9**: 3328–3344. doi: 10.1093/gbe/evx248
- Grover, C.E., Arick, M.A., Thrash, A., Conover, J.L., Sanders, W.S., Peterson, D.G., Frelichowski, J.E., Scheffler, J.A., Scheffler, B.E., and Wendel, J.F. 2019. Insights into the evolution of the New World diploid cottons (*Gossypium*, subgenus *Houzingenia*) based on genome sequencing. *Genome Biol. Evol.* **11**: 53–71. doi: 10.1093/gbe/evy256
- Grover, C.E., Gallagher, J.P., Jareczek, J.J., Page, J.T., Udall, J.A., Gore, M.A., and Wendel, J.F. 2015a. Re-evaluating the phylogeny of allopolyploid *Gossypium* L. *Mol. Phylogenet. Evol.* **92**: 45–52. doi: 10.1016/j.ympev.2015.05.023
- Grover, C.E., Grupp, K.K., Wanzek, R.J., and Wendel, J.F. 2012. Assessing the monophyly of polyploid *Gossypium* species. *Plant Syst. Evol.* **298**: 1177–1183. doi: 10.1007/s00606-012-0615-7
- Grover, C.E., Zhu, X., Grupp, K.K., Jareczek, J.J., Gallagher, J.P., Szadkowski, E., Seijo, J.G., and Wendel, J.F. 2015b. Molecular confirmation of species status for the allopolyploid cotton species, *Gossypium ekmanianum* Wittmack. *Genet. Resour. Crop Evol.* **62**: 103–114. doi: 10.1007/s10722-014-0138-x
- Hancock, J.F. 2012. *Plant evolution and the origin of crop species*. Wallingford, Oxfordshire, UK ; Cambridge, MA: CABI.
- Harlan, J.R. 1971. Agricultural origins: centers and noncenters: agriculture may originate in discrete centers or evolve over vast areas without definable centers. *Science* **174**: 468–474. doi: 10.1126/science.174.4008.468
- Harland, S.C. 1936. The genetical conception of the species. *Biol. Rev.* **11**: 83–112. doi: 10.1111/j.1469-185X.1936.tb00498.x
- Hawkins, J.S., Proulx, S.R., Rapp, R.A., and Wendel, J.F. 2009. Rapid DNA loss as a counterbalance to genome expansion through retrotransposon proliferation in plants. *Proc. Natl. Acad. Sci.* **106**: 17811–17816. doi: 10.1073/pnas.0904339106
- Hendrix, B. 2005. Estimation of the nuclear DNA content of *Gossypium* species. *Ann. Bot.* **95**: 789–797. doi: 10.1093/aob/mci078
- Hernández-Gutiérrez, R., and Magallón, S. 2019. The timing of Malvales evolution: incorporating its extensive fossil record to inform about lineage diversification. *Mol. Phylogenet. Evol.* **140**: 106606. doi: 10.1016/j.ympev.2019.106606
- Hinze, L.L., Gazave, E., Gore, M.A., Fang, D.D., Scheffler, B.E., Yu, J.Z., Jones, D.C., Frelichowski, J., and Percy, R.G. 2016. Genetic diversity of the two commercial tetraploid cotton species in the *Gossypium* diversity reference set. *J. Hered.* **107**: 274–286. doi: 10.1093/jhered/esw004
- Hu, G., Grover, C.E., Jareczek, J., Yuan, D., Dong, Y., Miller, E., Conover, J.L., and Wendel, J.F. 2021. Evolution and diversity of the cotton genome. In: Rahman, M., Y. Zafar, and T. Zhang, Eds., *Cotton precision breeding*. Cham: Springer International Publishing, pp. 25–78. doi: DOI: 10.1007/978-3-030-64504-5_2
- Hu, Y., Chen, J., Fang, L., Zhang, Z., Ma, W., Niu, Y., Ju, L., Deng, J., Zhao, T., Lian, J., et al. 2019. *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. *Nat. Genet.* **51**: 739–748. doi: 10.1038/s41588-019-0371-5
- Huang, G., Wu, Z., Percy, R.G., Bai, M., Li, Y., Frelichowski, J.E., Hu, J., Wang, K., Yu, J.Z., and Zhu, Y. 2020. Genome sequence of *Gossypium herbaceum* and genome updates of *Gossypium arboreum* and *Gossypium hirsutum* provide insights into cotton A-genome evolution. *Nat. Genet.* **52**: 516–524. doi: 10.1038/s41588-020-0607-4
- Huang, K., and Rieseberg, L.H. 2020. Frequency, Origins, and Evolutionary Role of Chromosomal inversions in plants. *Front. Plant Sci.* **11**: 296. doi: 10.3389/fpls.2020.00296
- Hutchinson, J.B. 1951. Intra-specific differentiation in *Gossypium hirsutum*. *Heredity* **5**: 161–193. doi: 10.1038/hdy.1951.19
- Hutchinson, J.B., Silow, R., and Stephens, S. 1947. *The evolution of Gossypium Gossypium and the differentiation of the cultivated cottons*. Empire Growing Corporation. Oxford University Press, London, UK, 160 p.
- Inomata, T., Triadan, D., Vázquez López, V.A., Fernandez-Diaz, J.C., Omori, T., Méndez Bauer, M.B., García Hernández, M., Beach, T., Cagnato, C., Aoyama, K., et al. 2020. Monumental architecture at Aguada Fénix and the rise of Maya civilization. *Nature* **582**: 530–533. doi: 10.1038/s41586-020-2343-

4

- Isendahl, C. 2011. The domestication and early spread of manioc (*Manihot esculenta* Crantz): a brief synthesis. *Lat. Am. Antiq.* **22**: 452–468. doi: 10.7183/1045-6635.22.4.452
- Jeanson N. 2020. Édition annotée et analyse d'une partie du catalogue de curiosités de la bibliothèque de Sainte-Geneviève par Claude du Molinet (Le Cabinet de la Bibliothèque de Sainte Geneviève, Paris, Dezallier, 1692) à savoir les quatre premiers chapitres de la seconde partie “ Histoire naturelle ”, ainsi que la préface générale, l'éloge du père du Molinet et les gravures synoptiques. *Sciences de l'Homme et Société*. <https://dumas.ccsd.cnrs.fr/dumas-02996017>
- Jenderek, M.M., and Frelichowski, J. 2019. Fiber crops: cotton and *Hesperaloe*. In: Greene, S.L., K.A. Williams, C.K. Khoury, M.B. Kantar, and L.F. Marek, Eds., *North American crop wild relatives, Volume 2*. Cham: Springer International Publishing, pp. 543–577. doi: 10.1007/978-3-319-97121-6_16
- Jiang, C., Chee, P.W., Draye, X., Morrell, P.L., Smith, C.W., and Paterson, A.H. 2000. Multilocus interactions restrict gene introgression in interspecific populations of polyploid *Gossypium* (cotton). *Evolution* **54**: 798–814. doi: 10.1111/j.0014-3820.2000.tb00081.x
- Jolie, E.A., Lynch, T.F., Geib, P.R., and Adovasio, J.M. 2011. Cordage, textiles, and the Late Pleistocene peopling of the Andes. *Curr. Anthropol.* **52**: 285–296. doi: 10.1086/659336
- Kennett, D.J., Lipson, M., Prufer, K.M., Mora-Marín, D., George, R.J., Rohland, N., Robinson, M., Trask, W.R., Edgar, H.H.J., Hill, E.C., et al. 2022. South-to-north migration preceded the advent of intensive farming in the Maya region. *Nat. Commun.* **13**: 1530. doi: 10.1038/s41467-022-29158-y
- Kennett, D.J., Prufer, K.M., Culleton, B.J., George, R.J., Robinson, M., Trask, W.R., Buckley, G.M., Moes, E., Kate, E.J., Harper, T.K., et al. 2020. Early isotopic evidence for maize as a staple grain in the Americas. *Sci. Adv.* **6**: eaba3245. doi: 10.1126/sciadv.aba3245
- Konan, N.O., Baudoin, J.-P., D'Hont, A., and Mergeai, G. 2009. Bridging Classical and Molecular Cytogenetics of *Gossypium*. In: Paterson, A.H., Ed., *Genetics and genomics of cotton*. New York, NY: Springer US, pp. 257–281. doi: 10.1007/978-0-387-70810-2_11
- Lazo, J. 1991. *Contribution à l'étude de l'origine, de la domestication et de la dispersion du cotonnier Gossypium barbadense L. : polymorphisme enzymatique, morphologique et agronomique des populations sauvages, subsponsantes et cultivées*. Thèse de doctorat en Sciences biologiques et fondamentales appliquées. Université Paris Sud Orsay. doi:
- Leitch, I.J., and Leitch, A.R. 2013. Genome size diversity and evolution in land plants. In: *Plant genome diversity*. Springer, Vienna, pp. 307–322.
- Lentz, D.L., Ramírez, C.R., and Griscom, B.W. 1997. Formative-period subsistence and forest-product extraction at the Yaramela Site, Honduras. *Anc. Mesoam.* **8**: 63–74. doi: 10.1017/S0956536100001577
- Lery, J. de 1580. *Histoire d'un voyage fait en la terre du Brésil, autrement dite Amérique*. Alphonse Lemerre, Éditeur, Paris.
- Lewton, F.L. 1920. The history of kidney cotton. *J. Wash. Acad. Sci.* 591–597.
- Li, F., Fan, G., Lu, C., Xiao, G., Zou, C., Kohel, R.J., Ma, Z., Shang, H., Ma, X., Wu, J., et al. 2015. Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nat. Biotechnol.* **33**: 524–530. doi: 10.1038/nbt.3208
- Linnaeus, C. 1753. *Species plantarum: exhibentes plantas rite cognitatas, ad genera relatas, cum differentiis specificis, nominibus trivialibus, synonymis selectis, locis natalibus, secundum systema sexuale digestas*. Holmiae: Impensis Laurentii Salvii, Stockholm, Sweden.
- Linnaeus, C. 1763. *Species plantarum: exhibentes plantas rite cognitatas, ad genera relatas, cum differentiis specificis, nominibus trivialibus, synonymis selectis, locis natalibus, secundum systema sexuale digestas*. 2nd edition. Impensis Laurentii Salvii, Stockholm.
- Liu, Q., Brubaker, C.L., Green, A.G., Marshall, D.R., Sharp, P.J., and Singh, S.P. 2001. Evolution of the FAD2-1 fatty acid desaturase 5' UTR intron and the molecular systematics of *Gossypium* (Malvaceae). *Am. J. Bot.* **88**: 92–102. doi: 10.2307/2657130
- Lohse, J.C., Awe, J., Griffith, C., Rosenswig, R.M., and Valdez, F. 2006. Preceramic occupations in Belize: updating the Paleoindian and Archaic record. *Lat. Am. Antiq.* **17**: 209–226. doi: 10.2307/25063047
- Lubbers, E.L., and Chee, P.W. 2009. The worldwide gene pool of *G. hirsutum* and its improvement. In:

- Paterson, A.H., Ed., *Genetics and genomics of cotton*. New York, NY: Springer US, pp. 23–52. doi: 10.1007/978-0-387-70810-2_2
- MacNeish, R.S. 1967a. An interdisciplinary approach to an archaeological problem. In: *The prehistory of the Tehuacan Valley, Volume One, Environment and subsistence*. Robert S. Peabody Foundation Phillips Academy Andover, University of Texas Press, Austin & London., pp. 14–24.
- MacNeish, R.S. 1967b. Introduction. In: *The Prehistory of the Tehuacan Valley, Vol.1: Environment and Subsistence*. University of Texas Press, Austin, TX, USA, pp. 3–12.
- MacNeish, R.S., Nelken-Terner, A., and Johnson, I.W. 1967. *The Prehistory of the Tehuacán Valley. Vol. 2: The non-ceramic artifacts*. University of Texas Press, Austin, TX, USA.
- Magallón, S., and Castillo, A. 2009. Angiosperm diversification through time. *Am. J. Bot.* **96**: 349–365. doi: 10.3732/ajb.0800060
- Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L.L., and Hernández-Hernández, T. 2015. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytol.* **207**: 437–453. doi: 10.1111/nph.13264
- Menzel, M.Y., and Brown, M.S. 1954. The Significance of multivalent formation in three-species *Gossypium* Hybrids. *Genetics* **39**: 546–557. doi: 10.1093/genetics/39.4.546
- Miksicek, C.A. 1990. Early Wetland Agriculture in the Maya Lowlands: Clues from Preserved Plant Remains In: *Ancient Maya wetland Agriculture: excavations on Albion Island, northern Belize*. Westview Press, Boulder, CO, USA, pp. 295–312.
- Molinet, C. du 1692. *Le Cabinet de la Bibliothèque de Sainte Geneviève*. Dezallier, Paris.
- Moret B.M., Nakhleh L., Warnow T., Linder C.R., Tholse A., Padolina A., Sun J., and Timme R. 2004. Phylogenetic networks: modeling, reconstructibility, and accuracy. *IEEE/ACM Trans Comput Biol Bioinform.* **1**:13-23. doi: 10.1109/TCBB.2004.10
- Moseley, M.E. 1992. Maritime foundations and multilineal evolution: retrospect and prospect. *Andean Past* **3**:5–42.
- Moulherat, C., Tengberg, M., Haquet, J.-F., and Mille, B. 2002. First Evidence of cotton at Neolithic Mehrgarh, Pakistan: Analysis of Mineralized Fibres from a Copper Bead. *J. Archaeol. Sci.* **29**: 1393–1401. doi: 10.1006/jasc.2001.0779
- Pagán-Jiménez, J.R., Rodríguez-Ramos, R., Reid, B.A., Bel, M. van den, and Hofman, C.L. 2015. Early dispersals of maize and other food plants into the southern Caribbean and northeastern South America. *Quat. Sci. Rev.* **123**: 231–246. doi: 10.1016/j.quascirev.2015.07.005
- Page, J.T., Huynh, M.D., Liechty, Z.S., Grupp, K., Stelly, D., Hulse, A.M., Ashrafi, H., Van Deynze, A., Wendel, J.F., and Udall, J.A. 2013. Insights into the evolution of cotton diploids and polyploids from whole-genome re-sequencing. *G3 Genes Genomes Genet.* **3**: 1809–1818. doi: 10.1534/g3.113.007229
- Pan, Y., Meng, F., and Wang, X. 2020. Sequencing multiple cotton genomes reveals complex structures and lays foundation for breeding. *Front. Plant Sci.* **11**: 560096. doi: 10.3389/fpls.2020.560096
- Pearsall, D.M. 2008. Plant domestication and the shift to agriculture in the Andes. In: Silverman, H., and W.H. Isbell, Eds., *The handbook of South American archaeology*. New York, NY: Springer New York, pp. 105–120. doi: 10.1007/978-0-387-74907-5_7
- Percival, A.E., Wendel, J.F., and Stewart, J.McD. 1999. Taxonomy and germplasm resources. In: Smith, C.W., and J.T. Cothren, Eds., *Cotton: origin, history, technology and production*. John Wiley & Sons, New York, pp. 33–63.
- Percy, R.G. 2009. The worldwide gene pool of *Gossypium barbadense* L. and its improvement. In: Paterson, A.H., Ed., *Genetics and genomics of cotton*. New York, NY: Springer US, pp. 53–68. doi: DOI: 10.1007/978-0-387-70810-2_3
- Percy, R.G., and Wendel, J.F. 1990. Allozyme evidence for the origin and diversification of *Gossypium barbadense* L. *Theor. Appl Genet.* **79**: 529–542. doi: 10.1007/BF00226164
- Phillips, L.L. 1963. The cytogenetics of *Gossypium* and the origin of New World cottons. *Evolution* **17**: 460. doi: 10.2307/2407096
- Phillips, L.L. 1966. The cytology and phylogenetics of the diploid species of *Gossypium*. *Am. J. Bot.* **53**: 328–335. doi: 10.1002/j.1537-2197.1966.tb07343.x
- Phillips, L.L. 1976. Cotton. In: Simmonds, N.W., *Evolution of crop plants*. Longman New York., pp. 196–200.

- Phuphathanaphong, L. 2006. *Thepparatia* (Malvaceae), a new genus from Thailand. *Thai For. Bull. Bot.* **34**: 195–200.
- Piperno, D.R. 2011. The origins of plant cultivation and domestication in the New World Tropics: patterns, process, and new developments. *Curr. Anthropol.* **52**: S453–S470. doi: 10.1086/659998
- Pool, C. 2007. *Olmec archaeology and Early Mesoamerica*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139167147
- Pope, K.O., Pohl, M.E.D., Jones, J.G., Lentz, D.L., Nagy, C. von, Vega, F.J., and Quitmyer, I.R. 2001. Origin and environmental setting of ancient agriculture in the lowlands of Mesoamerica. *Science* **292**: 1370–1373. doi: 10.1126/science.292.5520.1370
- Postel, Z., and Touzet, P. 2020. Cytonuclear genetic incompatibilities in plant speciation. *Plants* **9**: 487. doi: 10.3390/plants9040487
- Rapp, R.A., Alvarez, I., and Wendel, J.F. 2005. Molecular confirmation of the position of *Gossypium trifurcatum* Vollesen. *Genet. Resour. Crop Evol.* **52**: 749–753. doi: 10.1007/s10722-004-6097-x
- Rendón-Anaya, M., Montero-Vargas, J.M., Saburido-Álvarez, S., Vlasova, A., Capella-Gutierrez, S., Ordaz-Ortiz, J.J., Aguilar, O.M., Vianello-Brondani, R.P., Santalla, M., Delaye, L., et al. 2017. Genomic history of the origin and domestication of common bean unveils its closest sister species. *Genome Biol.* **18**: 60. doi: 10.1186/s13059-017-1190-6
- Reynaud, C., Thoury, M., Dazzi, A., Latour, G., Scheel, M., Li, J., Thomas, A., Moulhérat, C., Didier, A., and Bertrand, L. 2020. In-place molecular preservation of cellulose in 5,000-year-old archaeological textiles. *Proc. Natl. Acad. Sci.* **117**: 19670–19676. doi: 10.1073/pnas.2004139117
- Richardson, J.E., Whitlock, B.A., Meerow, A.W., and Madriñán, S. 2015. The age of chocolate: a diversification history of *Theobroma* and Malvaceae. *Front. Ecol. Evol.* **3**: 120. doi: DOI: 10.3389/fevo.2015.00120
- Rieseberg, L., and Soltis, D. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evol. Trends Plants* **5**(1): 65–84.
- Robert, S. 1997. *L'Egypte, passion française*. Editions du Seuil, Paris, France. ISBN 2-02-028144-9.
- Seelanan, T., Brubaker, C.L., Stewart, J.McD., Craven, L.A., and Wendel, J.F. 1999. Molecular systematics of Australian *Gossypium* Section *Grandicalyx* (Malvaceae). *Syst. Bot.* **24**: 183. doi: 10.2307/2419548
- Seelanan, T., Schnabel, A., and Wendel, J.F. 1997. Congruence and consensus in the cotton tribe (Malvaceae). *Syst. Bot.* **22**: 259. doi: 10.2307/2419457
- Senchina, D.S., Alvarez, I., Cronn, R.C., Liu, B., Rong, J., Noyes, R.D., Paterson, A.H., Wing, R.A., Wilkins, T.A., and Wendel, J.F. 2003. Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. *Mol. Biol. Evol.* **20**: 633–643. doi: 10.1093/molbev/msg065
- Silva, E.F. da, Duarte Filho, L.S.C., Santos Maia, A.K. dos, Silva, A.V. da, and Lins e Silva, A.C.B. 2018. The occurrence of wild cotton *Gossypium mustelinum* Miers ex Watt populations in the coast of Pernambuco, northeastern Brazil. *Genet. Resour. Crop Evol.* **65**: 1551–1557. doi: 10.1007/s10722-018-0637-2
- Skovsted, A. 1934. Cytological studies in cotton - II. Two interspecific hybrids between Asiatic and New World cottons. *J. Genet.* **28**: 407–424.
- Smith, B.D. 1997. The initial domestication of *Cucurbita pepo* in the Americas 10,000 years ago. *Science* **276**: 932–934. doi: 10.1126/science.276.5314.932
- Smith, C.E. 1967. Plant Remains In: Byers D. S., *The prehistory of the Tehuacan Valley, Vol.1: Environment and subsistence*. University of Texas Press, Austin, TX, USA, pp. 220–255.
- Smith, C.E., and Stephens, S.G. 1971. Critical identification of Mexican archaeological cotton remains. *Econ. Bot.* **25**: 160–168. doi: 10.1007/BF02860076
- Smith, C.W., and Cantrell, R.G. 1999. History of cultivar development in the United States In: Smith, C.W., and J.T. Cothren, Eds., *Cotton: origin, history, technology, and production*. John Wiley & Sons, New York, pp. 99–171.
- Söchtig, J., Álvarez-Iglesias, V., Mosquera-Miguel, A., Gelabert-Besada, M., Gómez-Carballa, A., and Salas, A. 2015. Genomic insights on the ethno-history of the Maya and the 'Ladinos' from Guatemala. *BMC Genomics* **16**: 131. doi: 10.1186/s12864-015-1339-1
- Splitstoser, J.C., Dillehay, T.D., Wouters, J., and Claro, A. 2016. Early pre-Hispanic use of indigo blue in Peru. *Sci. Adv.* **2**: e1501623. doi: 10.1126/sciadv.1501623

- Stark, B.L., Heller, L., and Ohnerson, M.A. 1998. People with cloth: mesoamerican economic change from the perspective of cotton in south-central Veracruz. *Lat. Am. Antiq.* **9**: 7–36. doi: 10.2307/972126
- Stephens, S.G. 1949. The cytogenetics of speciation in *Gossypium*. I. Selective elimination of the donor parent genotype in interspecific backcrosses. *Genetics* **34**: 627.
- Stephens, S.G. 1958. Salt water tolerance of seeds of *Gossypium* species as a possible factor in seed dispersal. *Am. Nat.* **92**: 83–92.
- Stephens, S.G. 1966. Problems on the origin, dispersal, and establishment of the Galapagos cottons. In: Bowman, R.I., Ed., *The Galápagos*. Proceedings of the Symposia of the Galápagos International Scientific Project. Los Angeles: University of California Press, pp. 201–208.
- Stephens, S.G. 1967. Evolution under domestication of the New World cottons (*Gossypium* spp.). *Cienc. E Cult.* 118–134.
- Stephens, S.G. 1973. Geographical distribution of cultivated cottons relative to probable centers of domestication in the New World. In: Srb, A.M., Ed., *Genes, enzymes, and populations*. Boston, MA: Springer US, pp. 239–254. doi:10.1007/978-1-4684-2880-3_16
- Stephens, S.G. 1975. A reexamination of the cotton remains from Huaca Prieta, north coastal Peru. *Am. Antiq.* **40**(4): 406–419.
- Stephens, S.G. 1976. The origin of Sea Island cotton. *Agric. Hist.* **50**(2): 391–399.
- Stephens, S.G., and Moseley, M.E. 1974. Early domesticated cottons from archaeological sites in central coastal Peru. *Am. Antiq.* **39**: 109–122. doi: 10.2307/279225
- Stephens, S.G., and Phillips, L.L. 1972. The history and geographical distribution of a polymorphic system in New World cottons. *Biotropica* **4**: 49. doi: 10.2307/2989727
- Stitzer, M.C., and Ross-Ibarra, J. 2018. Maize domestication and gene interaction. *New Phytol.* **220**: 395–408. doi: 10.1111/nph.15350
- Turcotte, E.L., and Percy, R.G. 1990. Genetics of Kidney Seed in *Gossypium barbadense* L. *Crop Sci.* **30**: 384. doi: 10.2135/cropsci1990.0011183X003000020029x
- Ulloa, M. 2014. The diploid D genome cottons (*Gossypium* spp.) of the New World In: Abdurakhmonov, I.Y., Ed., *World cotton germplasm resources*. InTech, Chapter 8. doi: 10.5772/58387
- Ulloa, M., Stewart, J.McD., Garcia-C., E.A., Godoy-A., S., Gaytan-M., A., and Acosta, N.S. 2006. Cotton genetic resources in the western states of Mexico: in situ conservation status and germplasm collection for ex situ preservation. *Genet. Resour. Crop Evol.* **53**: 653–668. doi: 10.1007/s10722-004-2988-0
- Viot, C. 2019. Domestication and varietal diversification of Old World cultivated cottons (*Gossypium* sp.) in the Antiquity. *Rev. d'ethnoécologie* **2019**: 15. doi: 10.4000/ethnoecologie.4404
- Viot, C., Jacobs, J., Arioli, T., Llewellyn, D., Derijcker, R., Claverie, M., and Lacape, M. 2009. Non-independence between markers on homoeologous chromosomes in an interspecific allopolyploid cotton RILs population In: *International conference on polyploidy, hybridization and biodiversity (ICPHB 2009)*, May 17-20, 2009, Saint-Malo - France, Program and Abstracts. Saint Malo, FRANCE, p. 135.
- Wang, B., Liu, L., Zhang, D., Zhuang, Z., Guo, H., Qiao, X., Wei, L., Rong, J., May, O.L., Paterson, A.H., et al. 2016. A genetic map between *Gossypium hirsutum* and the Brazilian endemic *G. mustelinum* and its application to QTL mapping. *G3 Genes Genomes Genet.* **6**: 1673–1685. doi: 10.1534/g3.116.029116
- Wang, J.-H., Moore, M.J., Wang, H., Zhu, Z.-X., and Wang, H.-F. 2021. Plastome evolution and phylogenetic relationships among Malvaceae subfamilies. *Gene* **765**: 145103. doi: 10.1016/j.gene.2020.145103
- Wang, M., Tu, L., Yuan, D., Zhu, D., Shen, C., Li, J., Liu, F., Pei, L., Wang, P., Zhao, G., et al. 2019. Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*. *Nat. Genet.* **51**: 224–229. doi: 10.1038/s41588-018-0282-x
- Watt, G. 1907. *The wild and cultivated cotton plants of the world*. London: Longmans, Green & Co.
- Wegier, A., Piñeyro-Nelson, A., Alarcón, J., Gálvez-Mariscal, A., Álvarez-Buylla, E.R., and Piñero, D. 2011. Recent long-distance transgene flow into wild populations conforms to historical patterns of gene flow in cotton (*Gossypium hirsutum*) at its centre of origin: historical and recent gene flow in wild cotton. *Mol. Ecol.* **20**: 4182–4194. doi: 10.1111/j.1365-294X.2011.05258.x
- Wen, D., Yu, Y., Zhu, J., & Nakhleh, L. 2018. Inferring phylogenetic networks using PhyloNet. *Syst. biol.*

- 67(4): 735–740. doi: 10.1093/sysbio/syy015
- Wendel, J.F. 1989. New World tetraploid cottons contain Old World cytoplasm. *Proc. Natl. Acad. Sci.* **86**: 4132–4136. doi: 10.1073/pnas.86.11.4132
- Wendel, J.F., and Albert, V.A. 1992. Phylogenetics of the cotton genus (*Gossypium*): character-state weighted parsimony analysis of chloroplast-DNA restriction site data and its systematic and biogeographic implications. *Syst. Bot.* **17**: 115. doi: 10.2307/2419069
- Wendel, J.F., Brubaker, C., Alvarez, I., Cronn, R., and Stewart, J.McD. 2009. Evolution and natural history of the cotton genus. In: Paterson, A.H., Ed., *Genetics and genomics of cotton*. New York, NY: Springer US, pp. 3–22. doi: 10.1007/978-0-387-70810-2_1
- Wendel, J.F., Brubaker, C.L., and Percival, A.E. 1992. Genetic diversity in *Gossypium hirsutum* and the origin of upland cotton. *Am. J. Bot.* **79**: 1291–1310. doi: 10.1002/j.1537-2197.1992.tb13734.x
- Wendel, J.F., Brubaker, C.L., and Seelanan, T. 2010. The origin and evolution of *Gossypium*. In: Stewart, J.McD., D.M. Oosterhuis, J.J. Heitholt, and J.R. Mauney, Eds., *Physiology of cotton*. Dordrecht: Springer Netherlands, pp. 1–18. doi: 10.1007/978-90-481-3195-2_1
- Wendel, J.F., and Cronn, R.C. 2003. Polyploidy and the evolutionary history of cotton. *Adv. Agron.* **78**: 139–186. doi: 10.1016/S0065-2113(02)78004-8
- Wendel, J.F., Cronn, R.C., Spencer Johnston, J., and James Price, H. 2002. Feast and famine in plant genomes. *Genetica* **115**: 37–47. doi: 10.1023/A:1016020030189
- Wendel, J.F., and Grover, C.E. 2015. Taxonomy and evolution of the cotton genus, *Gossypium* In: Fang, D.D., and R.G. Percy, Eds., *Cotton*. 2nd. ed.. Agronomy Monographs 57. Madison, WI, USA: American Society of Agronomy, Inc., Crop Science Society of America, Inc., and Soil Science Society of America, Inc., pp. 25–44. doi: 10.2134/agronmonogr57.2013.0020
- Wendel, J.F., and Percy, R.G. 1990. Allozyme diversity and introgression in the Galapagos Islands endemic *Gossypium darwinii* and its relationship to continental *G. barbadense*. *Biochem. Syst. Ecol.* **18**: 517–528. doi: 10.1016/0305-1978(90)90123-W
- Wendel, J.F., Schnabel, A., and Seelanan, T. 1995. An unusual ribosomal DNA Sequence from *Gossypium gossypioides* reveals ancient, cryptic, intergenomic introgression. *Mol. Phylogenet. Evol.* **4**: 298–313. doi: 10.1006/mpev.1995.1027
- Wendel, J.F., Stewart, J.McD., and Rettig, J.H. 1991. Molecular evidence for homoploid reticulate evolution among Australian species of *Gossypium*. *Evolution* **45**: 694–711. doi: 10.1111/j.1558-5646.1991.tb04339.x
- Westengen, O.T., Huamán, Z., and Heun, M. 2005. Genetic diversity and geographic pattern in early South American cotton domestication. *Theor. Appl. Genet.* **110**: 392–402. doi: 10.1007/s00122-004-1850-2
- Wu, Y., Liu, F., Yang, D.-G., Li, W., Zhou, X.-J., Pei, X.-Y., Liu, Y.-G., He, K.-L., Zhang, W.-S., Ren, Z.-Y., et al. 2018. Comparative chloroplast genomics of *Gossypium* species: insights into repeat sequence variations and phylogeny. *Front. Plant Sci.* **9**: 376. doi: 10.3389/fpls.2018.00376
- Yang, Z., Ge, X., Yang, Z., Qin, W., Sun, G., Wang, Z., Li, Z., Liu, J., Wu, J., Wang, Y., et al. 2019. Extensive intraspecific gene order and gene structural variations in upland cotton cultivars. *Nat. Commun.* **10**: 2989. doi: 10.1038/s41467-019-10820-x
- Yuan, D., Grover, C.E., Hu, G., Pan, M., Miller, E.R., Conover, J.L., Hunt, S.P., Udall, J.A., and Wendel, J.F. 2021. Parallel and intertwining threads of domestication in allopolyploid cotton. *Adv. Sci.* **8**: 2003634. doi: 10.1002/advs.202003634
- Yu-xiang, W., Jin-hong, C., Qiu-ling, H., and Shui-jin, Z. 2013. Parental origin and genomic evolution of tetraploid *Gossypium* species by molecular marker and GISH analyses. *Caryologia* **66**: 368–374. doi: 10.1080/00087114.2013.857830
- Zaitsev, G.S. 1928. A contribution to the classification of the genus *Gossypium* L. *Bull Appl Bot Genet Plant Breed.* **5**(8): 39–65.
- Zarrillo, S., Gaikwad, N., Lanaud, C., Powis, T., Viot, C., Lesur, I., Fouet, O., Argout, X., Guichoux, E., Salin, F., et al. 2018. The use and domestication of *Theobroma cacao* during the mid-Holocene in the upper Amazon. *Nat. Ecol. Evol.* **2**: 1879–1888. doi: 10.1038/s41559-018-0697-x
- Zeder, M.A. 2015. Core questions in domestication research. *Proc. Natl. Acad. Sci.* **112**: 3191–3198. doi: 10.1073/pnas.1501711112
- Zhang, S., Cai, Y., Guo, J., Li, K., Peng, R., Liu, F., Roberts, J.A., Miao, Y., and Zhang, X. 2019. Genotyping-by-sequencing of *Gossypium hirsutum* races and cultivars uncovers novel patterns of

- genetic relationships and domestication footprints. *Evol. Bioinforma.* **15**: 117693431988994. doi: 10.1177/1176934319889948
- Zhao, X., Si, Y., Hanson, R.E., Crane, C.F., Price, H.J., Stelly, D.M., Wendel, J.F., and Paterson, A.H. 1998. Dispersed repetitive DNA has spread to new genomes since polyploid formation in cotton. *Genome Res.* **8**: 479–492. doi: 10.1101/gr.8.5.479
- Zizumbo-Villarreal, D., and Colunga-GarcíaMarín, P. 2010. Origin of agriculture and plant domestication in West Mesoamerica. *Genet. Resour. Crop Evol.* **57**: 813–825. doi: 10.1007/s10722-009-9521-4
- Zupok, A., Kozul, D., Schöttler, M.A., Niehörster, J., Garbsch, F., Liere, K., Fischer, A., Zoschke, R., Malinova, I., Bock, R., et al. 2021. A photosynthesis operon in the chloroplast genome drives speciation in evening primroses. *Plant Cell* **33**: 2583–2601. doi: 10.1093/plcell/koab155

Supplementary data

Table S1. Subgeneric classification and relationship to genome groups of the 52 accepted species of *Gossypium*. Numbers of species given in parentheses; «-» denotes absence of taxonomic subdivision. Data sources: ARS-GRIN; Fryxell (1976, 1979); Wendel *et al.* (2009); Grover *et al.* (2015b); Gallagher *et al.* (2017).

Denomination in GRIN-ARS	Subgenus	Section	Subsection	Genome group	Genome
<i>Gossypium herbaceum</i> L.	<i>Gossypium</i>	<i>Gossypium</i>	<i>Gossypium</i>	A	A ₁
<i>Gossypium arboreum</i> L.	<i>Gossypium</i>	<i>Gossypium</i>	<i>Gossypium</i>	A	A ₂
<i>Gossypium anomalum</i> Wawra	<i>Gossypium</i>	<i>Gossypium</i>	<i>Anomala</i>	B	B ₁
<i>Gossypium triphyllum</i> (Harv.) Hochr.	<i>Gossypium</i>	<i>Gossypium</i>	<i>Anomala</i>	B	B ₂
<i>Gossypium capitis-viridis</i> Mauer	<i>Gossypium</i>	<i>Gossypium</i>	<i>Anomala</i>	B	B ₃
<i>Gossypium trifurcatum</i> Vollesen	<i>Gossypium</i>	<i>Serrata</i>	-	B?E?	B?E?
<i>Gossypium stocksii</i> Mast.	<i>Gossypium</i>	<i>Gossypium</i>	<i>Pseudopambak</i>	E	E ₁
<i>Gossypium somalense</i> (Gürke) J. B. Hutch.	<i>Gossypium</i>	<i>Gossypium</i>	<i>Pseudopambak</i>	E	E ₂
<i>Gossypium areysianum</i> Deflers	<i>Gossypium</i>	<i>Gossypium</i>	<i>Pseudopambak</i>	E	E ₃
<i>Gossypium incanum</i> (O. Schwartz) Hilc.	<i>Gossypium</i>	<i>Gossypium</i>	<i>Pseudopambak</i>	E	E ₄
<i>Gossypium benadirens</i> Mattei	<i>Gossypium</i>	<i>Gossypium</i>	<i>Pseudopambak</i>	E?	E?
<i>Gossypium bricchettii</i> (Ulbr.) Vollesen	<i>Gossypium</i>	<i>Gossypium</i>	<i>Pseudopambak</i>	E?	E?
<i>Gossypium vollesenii</i> Fryxell	<i>Gossypium</i>	<i>Gossypium</i>	<i>Pseudopambak</i>	E?	E?
<i>Gossypium longicalyx</i> J. B. Hutch. & B. J. S. Lee	<i>Gossypium</i>	<i>Gossypium</i>	<i>Longiloba</i>	F	F ₁
<i>Gossypium thurberi</i> Tod.	<i>Houzingenia</i>	<i>Houzingenia</i>	<i>Houzingenia</i>	D	D ₁
<i>Gossypium armourianum</i> Kearney	<i>Houzingenia</i>	<i>Houzingenia</i>	<i>Caducibracteolata</i>	D	D ₂₋₁
<i>Gossypium harknessii</i> Brandegee	<i>Houzingenia</i>	<i>Houzingenia</i>	<i>Caducibracteolata</i>	D	D ₂₋₂
<i>Gossypium davidsonii</i> Kellogg	<i>Houzingenia</i>	<i>Houzingenia</i>	<i>Integrifolia</i>	D	D _{3-d}
<i>Gossypium klotzschianum</i> Andersson	<i>Houzingenia</i>	<i>Houzingenia</i>	<i>Integrifolia</i>	D	D _{3-k}
<i>Gossypium aridum</i> (Rose & Standl.) Skovst.	<i>Houzingenia</i>	<i>Erioxylum</i>	<i>Erioxylum</i>	D	D ₄
<i>Gossypium raimondii</i> Ulbr.	<i>Houzingenia</i>	<i>Erioxylum</i>	<i>Austroamericana</i>	D	D ₅
<i>Gossypium gossypoides</i> (Ulbr.) Standl.	<i>Houzingenia</i>	<i>Erioxylum</i>	<i>Selera</i>	D	D ₆
<i>Gossypium lobatum</i> Gentry	<i>Houzingenia</i>	<i>Erioxylum</i>	<i>Erioxylum</i>	D	D ₇
<i>Gossypium trilobum</i> (DC.) Skovst.	<i>Houzingenia</i>	<i>Houzingenia</i>	<i>Houzingenia</i>	D	D ₈
<i>Gossypium laxum</i> L. L. Phillips	<i>Houzingenia</i>	<i>Houzingenia</i>	<i>Houzingenia</i>	D	D ₉
<i>Gossypium turneri</i> Fryxell	<i>Houzingenia</i>	<i>Houzingenia</i>	<i>Caducibracteolata</i>	D	D ₁₀
<i>Gossypium schwendimanii</i> Fryxell & S. D. Koch	<i>Houzingenia</i>	<i>Erioxylum</i>	<i>Erioxylum</i>	D	D ₁₁
<i>Gossypium sturtianum</i> J. H. Willis	<i>Sturtia</i>	<i>Sturtia</i>	-	C	C ₁
<i>Gossypium robinsonii</i> F. Muell.	<i>Sturtia</i>	<i>Sturtia</i>	-	C	C ₂
<i>Gossypium australe</i> F. Muell.	<i>Sturtia</i>	<i>Hibiscoidea</i>	-	G	G ₂
<i>Gossypium nelsonii</i> Fryxell	<i>Sturtia</i>	<i>Hibiscoidea</i>	-	G	G ₃
<i>Gossypium bickii</i> Prokh.	<i>Sturtia</i>	<i>Hibiscoidea</i>	-	G	G ₁
<i>Gossypium anapoides</i> J. M. Stewart et al.	<i>Sturtia</i>	<i>Grandicalyx</i>	-	K	K ₁₂
<i>Gossypium costulatum</i> Tod.	<i>Sturtia</i>	<i>Grandicalyx</i>	-	K	K ₈
<i>Gossypium cunninghamii</i> Tod.	<i>Sturtia</i>	<i>Grandicalyx</i>	-	K	K ₉
<i>Gossypium enthyle</i> Fryxell et al.	<i>Sturtia</i>	<i>Grandicalyx</i>	-	K	K ₇
<i>Gossypium exiguum</i> Fryxell et al.	<i>Sturtia</i>	<i>Grandicalyx</i>	-	K	K ₁

<i>Gossypium londonderriense</i> Fryxell et al.	<i>Sturtia</i>	<i>Grandicalyx</i>	-	K	K ₆
<i>Gossypium marchantii</i> Fryxell et al.	<i>Sturtia</i>	<i>Grandicalyx</i>	-	K	K ₅
<i>Gossypium nobile</i> Fryxell et al.	<i>Sturtia</i>	<i>Grandicalyx</i>	-	K	K ₁₁
<i>Gossypium pilosum</i> Fryxell	<i>Sturtia</i>	<i>Grandicalyx</i>	-	K	K ₄
<i>Gossypium populifolium</i> (Benth.) F. Muell. ex Tod.	<i>Sturtia</i>	<i>Grandicalyx</i>	-	K	K ₃
<i>Gossypium pulchellum</i> (C. A. Gardner) Fryxell	<i>Sturtia</i>	<i>Grandicalyx</i>	-	K	K ₁₀
<i>Gossypium rotundifolium</i> Fryxell et al.	<i>Sturtia</i>	<i>Grandicalyx</i>	-	K	K ₂
<i>Gossypium hirsutum</i> L.	<i>Karpas</i>	-	-	AD	(AD) ₁
<i>Gossypium barbadense</i> L.	<i>Karpas</i>	-	-	AD	(AD) ₂
<i>Gossypium tomentosum</i> Nutt. ex Seem.	<i>Karpas</i>	-	-	AD	(AD) ₃
<i>Gossypium mustelinum</i> Miers ex G. Watt	<i>Karpas</i>	-	-	AD	(AD) ₄
<i>Gossypium darwinii</i> G. Watt	<i>Karpas</i>	-	-	AD	(AD) ₅
<i>Gossypium ekmanianum</i> Wittm.	<i>Karpas</i>	-	-	AD	(AD) ₆
<i>Gossypium stephensii</i> J. Gallagher et al.	<i>Karpas</i>	-	-	AD	(AD) ₇

Table S2. Accepted subspecies and taxonomic varieties. Sources: id. Table S1.

Species name	Accepted infraspecific names	Genome group	Genome
<i>Gossypium anomalum</i> Wawra & Peyritsch	<i>Gossypium anomalum</i> subsp. <i>senarense</i> (Fenzl ex Wawra) Vollesen	B	B ₁
<i>Gossypium barbadense</i> L.	<i>Gossypium barbadense</i> var. <i>brasiliense</i> (G. Watt) J. B. Hutch. ex S. C. Harland	AD	(AD) ₂
<i>Gossypium herbaceum</i> L.	<i>Gossypium herbaceum</i> subsp. <i>africanum</i> (G. Watt) Vollesen	A	A _{1-a}
<i>Gossypium sturtianum</i> J. H. Willis	<i>Gossypium sturtianum</i> var. <i>nandewarensis</i> Derera	C	C _{1-n}