Chapter 2
Evolution and Diversity of the Cotton Genome

Guanjing Hu, Corrinne E. Grover, Daojun Yuan, Yating Dong, Emma Miller, Justin L. Conover, and Jonathan F. Wendel

2.1 Introduction

The cotton genus includes agronomically important species as well as many others that serve as examples of the evolution of biodiversity. Worldwide, cotton is most famous for the epidermal seed trichomes, or "fiber," of the cultivated species, the production of which comprises a multibillion dollar industry employing millions of people. Biologically, this crop is represented by four independently domesticated species at two different ploidy levels, generating additional interest as a naturally replicated evolutionary experiment. Accordingly, considerable attention has been given to the evolutionary relationships among species, the consequences of polyploidy, the domestication process, and the underlying biology that makes cotton a valuable crop species. In addition, recent technological advances continue to accelerate our understanding of cotton biology and evolution. Here we explore our understanding of cotton evolution and diversity, drawing attention both to the extraordinary evolutionary history of the genus and the importance of this diversity for agronomic improvement.

G. Hu
State Key Laboratory of Cotton Biology, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang, China

Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China

C. E. Grover · D. Yuan · Y. Dong · E. Miller · J. L. Conover · J. F. Wendel (✉)
Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA, USA
e-mail: jfw@iastate.edu

Mehboob-ur-Rahman et al. (eds.), *Cotton Precision Breeding*,
https://doi.org/10.1007/978-3-030-64504-5_2

2.2 Origin of the Cotton Genus (*Gossypium*)

Gossypium belongs to a small taxonomic tribe Gossypieae, both of which are characterized by the punctae or lysigenous cavities ("gossypol glands") that differentiate these taxa from other members of the family Malvaceae. The tribe Gossypieae is monophyletic (LaDuke and Doebley 1995; Seelanan et al. 1999; Wendel et al. 2002) and is divided into 9 small- to modestly sized genera which collectively include approximately 120 species (Fryxell 1968, 1979; Phuphathanaphong 2006). Five of these genera either are monotypic or contain fewer than four extant species, all of which have restricted geographic distributions: *Lebronnecia* (monotypic; Marquesas Islands), *Cephalohibiscus* (monotypic; New Guinea and the Solomon Islands), *Thepparatia* (monotypic; Thailand), *Gossypioides* (two species; East Africa and Madagascar), and *Kokia* (2–3 extant species, one extinct; Hawaii). The remaining genera are moderately sized with broader geographic ranges, i.e., *Hampea* (21 neotropical species), *Cienfuegosia* (25 species; neotropics and parts of Africa), *Thespesia* (17 tropical species), and *Gossypium* (Fig. 2.1).

Gossypium is the largest and most widely distributed genus in the tribe (Fryxell 1992), with over 55 recognized species whose naturally occurring ranges encompass the tropics and subtropics worldwide (Table 2.1). Despite their extensive distribution and extraordinary morphological and cytogenetic diversity, molecular phylogenetic analyses have confirmed the monophyletic origin of the cotton genus (Seelanan et al. 1999; Cronn et al. 2002) and identified its closest relative, i.e., the sister lineages *Kokia* and *Gossypioides*. This observation is somewhat surprising given the distant, restricted ranges of *Kokia* (Hawaii) and *Gossypioides* (East Africa, Madagascar) and their reduced chromosome number (n = 12), which is distinct from diploid species in *Gossypium* and most of the remainder of the cotton tribe (n = 13). That two genera now separated by over 17,500 km of open ocean are each other's closest relatives implies that transoceanic dispersal was involved in the evolution of one or both genera, representing one of the many examples of long-distance, salt-water dispersal in the tribe (Stephens 1958, 1966; Fryxell 1979; Wendel and Percival 1990; Percy and Wendel 1990; Wendel and Albert 1992; DeJoode and Wendel 1992). The distribution of the tribe itself includes many distant islands (e.g., Hawaii, Wake Island, Solomon Islands), whose phylogenetic distribution serves to underscore the importance of oceanic dispersal during the evolution of the tribe (Seelanan et al. 1999).

While the center for divergence of *Gossypium* consequently remains unclear (Grover et al. 2017a), the evolutionary timetable now is generally agreed upon (Senchina et al. 2003; Wendel and Grover 2015; Chen et al. 2016, 2017c, 2020). Divergence of *Gossypium* from its sister lineage, i.e., *Gossypioides-Kokia*, was initially estimated at approximately 10–15 million years ago (mya) based on relatively few chloroplast and nuclear genes (Seelanan et al. 1999; Cronn et al. 2002). This estimate was more recently upheld using de novo whole-genome sequences and thousands of genes representing the entire genome (Grover et al. 2017a). Interestingly, divergence between *Kokia* and *Gossypioides* was estimated at approximately 3–5

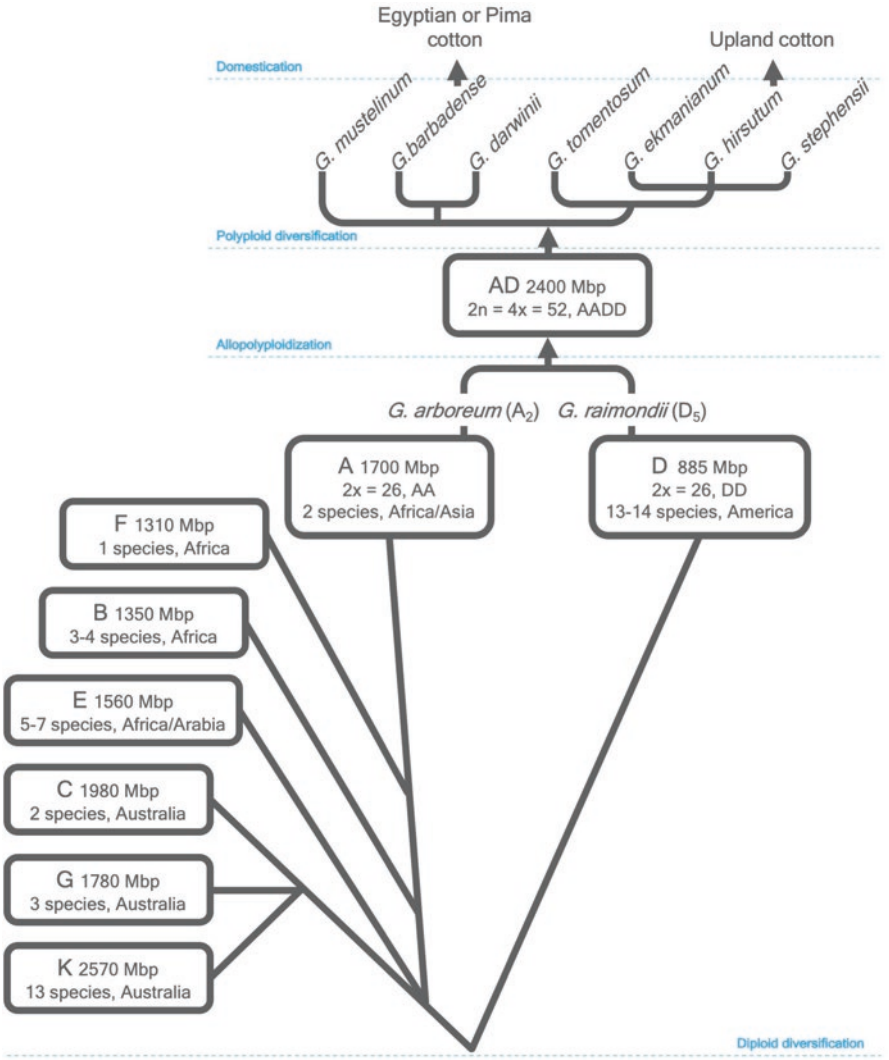

Fig. 2.1 Evolutionary history of *Gossypium.* Following the genus origin 5–10 Mya, diploid *Gossypium* rapidly diversified into three major lineages of eight monophyletic genome groups: the New World clade (D), the African-Asian clade (A, B, E and F), and the Australian clade (C. G and K). This global radiation involved several trans-oceanic dispersal events and was accompanied by morphological, ecological, and genome size differentiation. During the Pleistocene 1–2 mya, allopolyploid cottons formed following trans-oceanic dispersal of an A-genome diploid to the Americas, where the immigrant underwent hybridization, as female, with a native D-genome diploid similar to modern *G. raimondii*. Among the seven modern allopolyploid species, *G hirsutum* and *G. barbadense* were independently domesticated for fiber production

Table 2.1 Taxonomy of *Gossypium* species and notable features

Gossypium L.	Genome designations	Species presently recognized		Comments	Reference
Subgenus *Gossypium*				African, Arabian, and Asian diploid species	
Section *Gossypium*				Wild forms known only from Southern Africa, but with an indigenous domesticated range encompassing parts of Africa, the Middle East, and Asia	Fryxell (1979, 1992); Vollesen (1987)
Subsection *Gossypium*			2	The female parent of the allopolyploid formation was an A-genome progenitor	Wendel et al. (1989)
	A_1	*G. herbaceum* Linnaeus		Type species of the genus. Cultivated on a small scale but is used as a germplasm pool for many desirable traits: bacterial blight resistance, *R. reniformis* resistance, thrip resistance	Endrizzi et al. (1985); Stewart (1995)
	A_2	*G. arboreum* Linnaeus		Cultivated on a small scale but is used as a germplasm pool for many desirable traits: *R. reniformis* resistance, thrip resistance	Endrizzi et al. (1985); Stewart (1995)
Subsection *Anomala* Todaro			3–4	Africa and Cape Verde Islands	
	B_1	*G. anomalum* Wawra and Peyritsch		Germplasm pool for bacterial blight resistance, fiber improvement (finer and stronger), and cotton rust resistance	Endrizzi et al. (1985); Mehetre (2010); Fryxell et al. (1984)
	B_2	*G. triphyllum* (Harvey and Sonder) Hochreutiner		Trifoliolate leaf with fine tomentum that limits water loss from transpiration. Adapted to an extreme desert environment. Molecular data suggests affinities are with other B-genome species	Wendel and Albert (1992); Seelanan et al. (1997); Fryxell (1986)
	B_3	*G. capitis-viridis* Mauer		From Cape Verde Islands	Vollesen (1987); Fryxell et al. (1984)
	B*	(*G. trifurcatum*)			Vollesen (1987)

(continued)

Table 2.1 (continued)

Gossypium L.	Genome designations	Species presently recognized		Comments	Reference
Subsection *Pseudopambak* (Prokhanov) Fryxell			5–7	Poorly understood native ranges of many species, but collectively in parts of the Arabian Peninsula, Northeast Africa, and Southwest Asia	Vollesen (1987); Fryxell (1992)
	E_1	*G. stocksii* Masters in Hooker		Potential source of cotton leaf curl resistance and improved fiber strength. Demonstrates increase resistance to *R. reniformis*	Nazeer et al. (2014); Yik and Birchfield (1984)
	E_2	*G. somalense* (Gurke) Hutchinson		Demonstrates increase resistance to *R. reniformis*. Potential complementary lethal factor when bred with *G. hirsutum*	Yik and Birchfield (1984)
	E_3	*G. areysianum* Deflers		Foliage smells putrid	
	E_4	*G. incanum* (Schwartz) Hillcoat		Tends to grow in dry beds that flash flood	Fryxell (1986)
	E*	(*G. benadirense* Mattei) (*G. bricchettii* (Ulbrich) Vollesen) (*G. vollesenii* Fryxell)		Reinstated from limited materials thus incompletely understood	Vollesen (1987); Fryxell (1992)
Subsection *Longiloba*			1		
	F_1	*G. longicalyx* Hutchinson and Lee		Endemic to East Africa. Cytologically unique and unusually adapted for mesic environments preferring shade and higher rainfall. Useful in breeding schemes that transfer desirable wild traits to cultivated accessions. Like the A-genome, it has long fibers. Completely immune to *R. reniformis*	Phillips and Strickland (1966); Fryxell (1979, 1986); Wendel et al. (2010); Bell et al. (2014, 2015); Wendel and Grover (2015); Yik and Birchfield (1984)

(continued)

Table 2.1 (continued)

Gossypium L.	Genome designations	Species presently recognized		Comments	Reference
Subgenus *Houzingenia* (Fryxell) Fryxell			13–14	New-World diploids; large shrubs and small trees in the South and Middle America (primarily Mexico), with range extensions into Peru, the Galapagos Islands, and Southern Arizona	
Section *Houzingenia*				Large shrubs and small trees found primarily in Mexico	Fryxell 1(992)
Subsection *Houzingenia*				The two species in this subsection are morphologically similar and interfertile	Fryxell (1965, 1968, 1979)
	D_1	*G. thurberi* Todaro		Northernmost species adapted to temperate climate into the mountains of Arizona, which tolerates mild frost via defoliation and becomes fully dormant during the winter. D-genome species employed by J.O. Beasley to create the triple hybrid that was used to introgress high fiber strength into *G. hirsutum.* Tends to grow in dry beds that flash flood. Resistance against thirps and silverleaf whitefly which carries disease such as cotton leaf crumple. Lacks seed fiber	Fryxell (1979, 1986); Fryxell et al. (1984); Walker and Natwick (2006)
	D_8	*G. trilobum* (DC.) Skovsted		Fully tropical and susceptible to freezing temperature; grown in higher elevation than others (up to 2600 m) and perhaps the least xerophytic in distribution. Source of male sterile cytoplasm and restorer factor resistance to thrips	Fryxell (1965, 1968); Fryxell et al. (1984); Stewart (1992)

(continued)

Table 2.1 (continued)

Gossypium L.	Genome designations	Species presently recognized		Comments	Reference
Subsection *Integrifolia* (Todaro) Todaro				Interspecific hybrids between either species in this subsection and several other species except B, C, and G are embryo lethal	Phillips (1977); Lee (1981, 1986)
D_{3d}		*G. davidsonii* Kellogg		Source of salt tolerance. Broad ecological range in Baja California, with primary distribution in the Cape Region up to 2000 ft in elevation from seaside; seaside locations tend to be back from the leading edge of salt spray zone	Wendel and Percival (1990); Zhang et al. (2016); J Nason (personal observation)
D_{3k}		*G. klotzschianum* Andersson		A derivative species of *G. davidsonii* from Baja California to the Galapagos Islands through long-distance dispersal. Source of salt tolerance	Wendel and Percival (1990); Wei et al. (2017)
Subsection *Caducibracteolata* Mauer				These species are calciphiles, typically found in arid habitats around the Gulf of California. They have reduced leaves with thick cuticles and a double palisade layer and the largest seeds of the diploid species. The floral bracts are caduceus, abscising well before anthesis in *G. armourianum*, and shortly before to just after anthesis in the other species. They are subjected to high temperatures, high isolation, and low rainfall almost year-round	Phillips and Clement (1967); Fryxell (1986, 1992)

(continued)

Table 2.1 (continued)

Gossypium L.	Genome designations	Species presently recognized		Comments	Reference
	$D_{2\text{-}1}$	*G. armourianum* Kearney		Germplasm pool for bacterial blight and *R. reniformis* resistance. Large seeds 8–10 mm. Adapted to extreme water loss with leaves that have a double layer of palisade cells and a thick cuticle	Endrizzi et al. (1985); Fryxell et al. (1984)
	$D_{2\text{-}2}$	*G. harknessii* Brandegee		Source of cytoplasmic male sterility and restorer factors. Large seeds 8–10 mm. Adapted to extreme water loss with leaves that have a double layer of palisade cells and a thick cuticle	Meyer (1975); Fryxell et al. (1984)
	D_{10}	*G. turneri* Fryxell		Sister species to or derivative from *G. harknessii*. Adapted to extreme water loss with leaves that have a double layer of palisade cells and a thick cuticle	Fryxell et al. (1984); Fryxell (1986)
Section *Erioxylum* (Rose and Standley) Prokhanov					
Subsection *Erioxylum*				This group of species has a unique flowering phenology adapted to wet season-dry season cycle. At the height of the dry season, while leafless, the plants flower and fruit. After the fruits mature, the plants remain dormant until returning rains to stimulate new vegetative growth	Fryxell et al. (1984)
	D_4	*G. aridum* (Rose and Standley) Skovsted		Widely distributed and abundant in western Mexico. There is evidence of cytoplasmic introgression from subsection *Integrifolia* into populations of *G. aridum* from the State of Colima, Mexico; other populations are normal in this respect. Source of salt tolerance	Wendel et al. (1995a, b); Fryxell (1979); Fan et al. (2015)

(continued)

Table 2.1 (continued)

Gossypium L.	Genome designations	Species presently recognized		Comments	Reference
D_7		*G. lobatum* Gentry		Restricted distribution while locally abundant in Michoacán, Mexico. Leaves nearly distichous	Fryxell (1979, 1992)
D_9		*G. laxum* Phillips		Restricted distribution in Guerrero, Mexico, while locally abundant as subdominant species in the vegetation of Cañón del Zopilote	DeJoode and Wendel (1992); Wendel and Albert (1992)
D_{11}		*G. schwendimanii* Fryxell and Koch		The most recently described species among the New-World diploids	Fryxell (1979)
D_{12}		(*G. sp.nov.*)		A newly collected accession, US-72, genetically distant from other species in the subsection	Feng et al. (2011)
Subsection *Selera* (Ulbrich) Fryxell					
D_6		*G. gossypioides* (Ulbrich) Standley		The only diploid species that shows evidence of the original A×D hybridization that gave rise to the allotetraploids. This species may have arisen via introgressive speciation. Lacks foliar nectaries	Fryxell et al. (1987); Fryxell et al. (1984)
Subsection *Austroamericana* Fryxell					
D_5		*G. raimondii* Ulbrich		A relatively recent immigrant to Peru, model of the D-genome parent of allopolyploid cotton. Tends to grow in dry beds that flash flood	Endrizzi et al. (1985); Wendel et al. (1995a, b); Fryxell (1986)
Subgenus *Sturtia* (R. Brown) Todaro				All the indigenous Australian species	Fryxell (1992)
Section *Sturtia*			2	Central and Western Australia. Mauve flower. Neither of these deposit gossypol. Useful in studying the regulation and biosynthesis of gossypol production. Breeding schemes have already introduced the "glandless-seed and glanded plant" phenotype to *G. hirsutum*	Fryxell (1992); Brubaker et al. (1996); Mammadov et al. (2018); Zhu et al. (2005); Liu et al. (2015b)

(continued)

Table 2.1 (continued)

Gossypium L.	Genome designations	Species presently recognized		Comments	Reference
	C_1	*G. sturtianum* Willis		"Sturt's Desert Rose," the floral emblem of the Northern Territory, is distributed widely across the Australian continent in the temperate arid zone. Limited cold resistant, it can withstand a few degrees below freezing when in full leaf. Waxy leaves prevent water loss and can fold leaves inward when exposed to water stress. Small seeds 4–5 mm	Fryxell (1979, 1986); Fryxell et al. (1984); Craven et al. (1994)
	C_2	*G. robinsonii* Mueller		Potentially basal in the Australian *Gossypium* lineage. Prefers to grow in intermittent water beds to get direct access to moisture for a brief period	Craven et al. (1994); Fryxell (1986)
Section *Grandicalyx* Fryxell			13	Northwest Australia (especially the Kimberley region), Cobourg Peninsula, and Northern Territory, Australia. In contrast to large shrubs of other Australian species, these are sub-shrubby and produce short-lived stems from a perennial rootstock, as an adaption to the dry-season fires of the monsoon zone. White flowers with strongly contrasting red petal spots, occasional pink flowers occur; seeds rely on ant dispersal. Additionally, this section has the largest genome size in all of *Gossypium*	Fryxell (1992); Fryxell (1992); Craven et al. (1994); Stewart (1995); Wendel et al. (2010)
	K_8	*G. costulatum* Todaro		One of the first Australian *Gossypium* species to be collected, along with *G. cunninghamii* and *G. populifolium*, by Alan Cunningham between 1818 and 1820 where each occurs near coastal waters in northwest Australia accessible by ship	Craven et al. (1994)

(continued)

Table 2.1 (continued)

Gossypium L.	Genome designations	Species presently recognized		Comments	Reference
K_9		*G. cunninghamii* Todaro		May have originated from a hybridization event. The only sessile or subsessile species in *Gossypium*. This species may have originated from an ancient hybridization in which one parent (maternal) was a species similar to present-day *G. sturtianum*. A similar cytoplasm is also found in *G. bickii*. The paternal parent, however, is located in the Northern Territory	Wendel and Albert (1992)
K_1		*G. exiguum* Fryxell, Craven and Stewart		More widely distributed than other species in this section. *G. exiguum*, *G. rotundifolium*, and *G. pilosum* may be difficult to distinguish in the field and may have imprecise taxonomic descriptions	Fryxell (1992); Stewart, Craven, Brubaker and Wendel (personal observations)
K_2		*G. rotundifolium* Fryxell, Craven and Stewart			
K_4		*G. pilosum* Fryxell			
K_3		*G. populifolium* (Bentham) Mueller ex Todaro		These suberect and erect species are more narrowly distributed than the former three species, with some being known from only a few populations	Fryxell (1992); Stewart, Craven, Brubaker and Wendel (personal observations)
K_5		*G. marchantii* Fryxell, Craven and Stewart			
K_6		*G. londonderriense* Fryxell, Craven and Stewart			
K_7		*G. enthyle* Fryxell, Craven and Stewart			
K_{10}		*G. pulchellum* (Gardner) Fryxell			
K_{11}		*G. nobile* Fryxell, Craven and Stewart			
K_{12}		*G. anapoides* Stewart, Craven and *Wendel*			

(continued)

Table 2.1 (continued)

Gossypium L.	Genome designations	Species presently recognized		Comments	Reference
Section *Hibiscoidea* Todaro			3	Central Australia. Like *Sturtia* they do not deposit gossypol in seeds. Recurved pedicels which could prevent boll rot and rain damage to lint	Brubaker et al. (1996); Brown and Ware (1958)
	G_1	*G. bickii* Prokhanov		Created by hybridization of a maternal *G. sturtianum*-like plant and a paternal *G. australe-/ nelsonii*-like plant	Wendel et al. (1992)
	G_2	*G. australe* Mueller		Wind dispersed. Small seeds 4–5 mm. Fibers are straight	Wendel et al. (2010); Fryxell et al. (1984)
	G_3	*G. nelsonii* Fryxell		Wind dispersed. Fibers are straight	Wendel et al. (2010); Fryxell et al. (1984)
Subgenus *Karpas*			7	The allopolyploid cottons were probably formed during the Pleistocene by hybridization of diploids from the A and D genomes	Wendel et al. (1989); Wendel and Albert (1992); Seelanan et al. (1997)
	AD_1	*G. hirsutum* Linnaeus		Large amount of phenotypic diversity which provides many agronomically desirable traits. Originally distributed in Central America	Bell (1984); Meredith (1991); Brubaker et al. (1993); Brubaker and Wendel (1994)
	AD_2	*G. barbadense* Linnaeus		Distributed and domesticated in South America. Modern cultivars are highly introgressed with *G. hirsutum.* Some accessions have increased resistance to thrips and increased fiber quality	Shepherd (1974); Meredith (1991); Percy and Wendel (1990); Wang et al. (1995); Zhang et al. (2013)
	AD_3	*G. tomentosum* Nuttall ex Seemann		Endemic to the Hawaiian Islands and lacks foliar nectaries. Resistance to thrips and verticillium. Important source of salt tolerance	Zhang et al. (2013); Meyer and Meyer (1961); DeJoode and Wendel et al. (1992); Zhang et al. (2011); Oluoch et al. (2016)

(continued)

Table 2.1 (continued)

Gossypium L.	Genome designations	Species presently recognized		Comments	Reference
AD_4		*G. mustelinum* Miers ex Watt		Widely scattered populations in NE Brazil; resistance to thrips	Wendel et al. (1994); Bowman and McCarty (1997)
AD_5		*G. darwinii* Watt		Closely related to *G. barbadense* and found on the Galapagos Islands. Source of resistance to *Fusarium* and *Verticillium*	Bell (1984); Wendel and Percival (1990); Liu et al. (2016)
AD_6		*G. ekmanianum* Wittmack		Suggested as a separate species in 1928 but not confirmed until 2014	Grover et al. (2014)
AD_7		*G. stephensii* J. Gallagher, C. Grover and Wendel		Inhabits Wake Atoll	Gallagher et al. (2017)

Citations in this table: Meyer and Meyer (1961); Fryxell (1965, 1979, 1986, 1992); Phillips and Strickland (1966); Phillips and Clement (1967); Shepherd (1974); Meyer (1975); Phillips (1977); Bell (1984); Fryxell et al. (1984, 1987); Yik and Birchfield (1984); Endrizzi et al. (1985); Vollesen (1987); Wendel et al. (1989); Wendel et al. (1989, 1994, 1995a, 2010); Wendel and Percival (1990); Percy and Wendel (1990); Meredith (1991); Stewart (1992, 1995); Wendel and Albert (1992); DeJoode and Wendel (1992); Brubaker et al. (1993, 1996); Brubaker and Wendel (1994); Craven et al. (1994); Wang et al. (1995); Bowman and McCarty (1997); Seelanan et al. (1997); Zhu et al. (2005); Walker and Natwick (2006); Mehetre (2010); Feng et al. (2011); Zhang et al. (2013, 2016); Nazeer et al. (2014); Bell et al. (2014, 2015); Grover et al. (2015c); Wendel and Grover (2015); Fan et al. (2015); Liu et al. (2015b); Oluoch et al. (2016); Gallagher et al. (2017); Wei et al. (2017); Mammadov et al. (2018)

mya (Grover et al. 2017a), which is approximately the same time during which the major lineages of *Gossypium* became established and began to diversify (Cronn et al. 2002; Grover et al. 2019a). During this time, the *Kokia-Gossypioides* lineages experienced a shared reduction in chromosome number (Seelanan et al. 1999; Udall et al. 2019) and subsequently experienced remarkable genome downsizing, including massive, differential gene loss (Grover et al. 2017a). This insight into the *Kokia-Gossypioides* lineage provides an essential context in using these genera as phylogenetic out-groups, having consequences for understanding genome evolution, as well as evolutionary patterns and processes within *Gossypium*.

2.3 Diversification of *Gossypium* Diploid Species

Diploid members of *Gossypium* are divided into eight monophyletic genome groups (A through G and K; Table 2.1; Fig. 2.1), as determined by interspecific meiotic pairing behavior (Beasley 1940, 1942; Endrizzi et al. 1985; Zhang and Endrizzi 2015). Classification within and among these groups reflects decades of accumulated understanding that emerged from basic plant exploration, as well as taxonomic

and evolutionary analysis (Watt 1907; Hutchinson et al. 1947; Saunders 1961; Fryxell 1979, 1992; Wendel et al. 2009; Wendel and Grover 2015; Wang et al. 2018a). According to the most modern and widely followed taxonomic classification of Fryxell (1979, 1992), species are grouped into four subgenera and eight sections primarily based on morphological, cytogenetic, and geographical evidence. Subsequent molecular phylogenetic analyses (Wendel and Albert 1992; Cronn et al. 1996; Seelanan et al. 1999; Feng et al. 2011; Grover et al. 2015b, c, 2019a; Chen et al. 2016; Gallagher et al. 2017) have confirmed that the genealogical lineages of *Gossypium* species are largely congruent with genome designations and geographic distributions. That is, each genome group is monophyletic, representing a single natural lineage, and in most cases, these lineages are also geographically cohesive. Corresponding to continental division, the three major lineages of diploid species include the Australian subgenus *Sturtia* (C-, G-, K-genomes), the American subgenus *Houzingenia* (D-genome) of the New World, and the African, Arabian, and Asian subgenus *Gossypium* (A-, B-, E-, and F-genomes) of the Old World. Phylogenetic analyses place the earliest divergence between the New World (D-genome) lineage and the ancestor of all African/Australian taxa, followed shortly by divergence of the Australian cottons (i.e., C-, G-, and K-genomes), although the internodes in this basalmost split are short and hence uncertain. Diversification continued with the divergence of the African-Arabian E-genome species from the African A-, B-, and F-genome cottons, and, finally, B-genome divergence from the sister clades composed of the A- and F-genomes, the latter having a single species, *G. longicalyx*. The observation that *G. longicalyx* is sister to the A-genome species is particularly important in that *G. longicalyx* represents the wild ancestor in the evolution of spinnable fiber, which is unique (among diploids) to the A-genome.

Morphologically, the genus *Gossypium* is quite diverse (Fig. 2.2). Cotton species have collectively achieved a nearly worldwide distribution, with species-rich regions

Fig. 2.2 Striking morphological diversity of *Gossypium.* Reprinted (adapted) with permission from Cotton Incorporated

in the arid or seasonally arid tropics and subtropics over all major continents (see Table 2.1 for geographic distribution of species). Consequently, and in response to ecological and environmental demands, plant habits range from the fire-adapted, herbaceous perennials in NW Australia to small trees in SW Mexico that escape the dry season by dropping their leaves. Corolla colors are equally variable, spanning a rainbow that includes blues/purples (e.g., *G. triphyllum*), mauves/pinks (e.g., "Sturt's Desert Rose," *G. sturtianum*), whites/pale yellows (multiple species from NW Australia, Mexico, Africa-Arabia), and even a deep sulfur-yellow (i.e., *G. tomentosum* from Hawaii). Seed coverings range from nearly glabrous (e.g., *G. klotzschianum* and *G. davidsonii*) to short stiff, dense, brown hairs that aid in wind dispersal (*G. australe* and *G. nelsonii*), to the long, fine white fibers that characterize highly improved forms of the four cultivated species. There are even seeds that produce fat bodies to facilitate ant dispersal (section *Grandicalyx* from NW Australia; Seelanan et al. 1999). At the other end of the ant coevolution spectrum is *G. tomentosum* from the Hawaiian Islands, which lost the foliar and extrafloral nectaries that are common in other *Gossypium* species, presumably in response to the absence of native ants. Much of this morphological diversity is described in detail by Fryxell (1979).

In addition to the impressive morphological and ecological diversification, extensive chromosomal evolution has been documented for the genus (Endrizzi et al. 1985; Konan et al. 2009; Soltis et al. 2009). As mentioned above, the genome groups were originally designated based on chromosome size and the meiotic behavior of chromosomes during interspecific crosses. That is, those within the same genome group generally produce hybrids exhibiting bivalent chromosome formation (Beasley 1940, 1942; Endrizzi et al. 1985; Zhang and Endrizzi 2015), and increased occurrence of univalent formation generally reflects structural differentiation between more distant species (Gerstel 1953; Phillips 1966). Notably, although all diploid species share $n = 13$, there is more than threefold variation in DNA content per cell (Hendrix and Stewart 2005) that is evident even from gross chromosome morphology (Stephens 1947; Katterman and Ergle 1970; Abdul Kadir 1976). Genome sizes range from less than 1 Gbp in the American D-genome cottons (average 900 Mbp) to over 2.5 Gbp in the Australian K-genome species (Wendel et al. 2002; Hendrix and Stewart 2005), with most diploid cottons ranging between 1300 and 2000 Mbp. This extraordinary variation in genome size is largely attributed to changes in repetitive DNA content, particularly LTR transposable elements (TEs) (Hawkins et al. 2006; Grover et al. 2008, 2017a, 2019a; Renny-Byfield et al. 2016). Comparisons among closely related species of similar genome size have also revealed a cryptic and dynamic scenario of genome size gain and loss. Recent comparative studies within the A- (Renny-Byfield et al. 2016) and D-genome groups (Grover et al. 2019a), as well as between the outgroup genera *Kokia* and *Gossypioides* (Grover et al. 2017a) whose genome size appears static at 520 Mbp, found that relative stasis in size belies a much more complicated scenario of opposing actions, i.e., TE proliferation, TE loss, and small-scale insertions and deletions. Recent estimates of gene numbers also suggest more variability in gene content than previously considered (Page et al. 2013; Grover et al. 2019a), although further research with high-quality genomes is required to characterize gene evolution in diploid cotton.

2.4 Origin and Diversification of Polyploids

Allopolyploid cotton is the fortuitous result of one of the transoceanic dispersals that characterize the Gossypieae. Over half a century ago, a rich body of classic cytogenetic evidence followed by numerous experimental studies established that cotton species with $n = 26$ are allopolyploids composed of two co-resident genomes (aka subgenomes), one from an African A-genome species and the other from a Mesoamerican D-genome species (reviewed in detail by Endrizzi et al. (1985) and Wendel and Cronn (2003)). Since the native distribution of most polyploid species is primarily Mesoamerican, the formation of the AD-genome allopolyploid species most likely occurred somewhere in Mesoamerica following the long-distance transoceanic dispersal of an African A-genome species to the New World. Given the presumed rarity of such transoceanic dispersals, formation of the allopolyploid (i.e., A-/D-genome hybridization and subsequent doubling) is inferred to have occurred only once, in contrast to the prevalence of multiple polyploid origins in other plant species (Soltis et al. 1993, 2004; Soltis and Soltis 1999; Tate et al. 2006). Indeed a monophyletic origin for the polyploid clade is further supported by molecular evidence (Grover et al. 2012a, 2015a), most recently using whole-genome sequences for five allopolyploid species and both model diploid progenitors (Chen et al. 2020). Although hypotheses for the emergence of the polyploid clade originally ranged from a very ancient Cretaceous origin, perhaps 60–100 mya (prior to the separation of the South American and African continents; Saunders 1961; Endrizzi et al. 1985, 1989), to a very recent human-mediated origin (circa 6000 years ago via intentional intercontinental transfer; Hutchinson et al. 1947; Hutchinson 1959; Johnson 1975), a mid-Pleistocene origin (1–2 mya) was generally supported by botanists based on cytogenetic evidence (Phillips 1964) and ecological considerations (Fryxell 1965) and subsequently confirmed by numerous sources of molecular and DNA sequencing data (Wendel and Cronn 2003; Wendel and Grover 2015).

Numerous sources of molecular and sequence data have also supported the identification of the closest model progenitors to the actual parents of the polyploid clade. *Gossypium raimondii* has long been recognized as the closest extant relative of the actual D-genome parent of allopolyploid cotton, from the earliest efforts comparing leaf development between diverse synthetic hybrids and the natural allopolyploids to modern sequencing data (Stephens 1944; Endrizzi et al. 1985; Wendel 1989; Wendel and Cronn 2003; Wendel et al. 2012; Yu-xiang et al. 2013; Li et al. 2014; Grover et al. 2015a). The A-genome parent was historically less clear, although some have considered *G. herbaceum* to be a better candidate than *G. arboreum* based on cytogenetics (Endrizzi et al. 1985). Subsequent molecular and phylogenetic evidence, however, supports sister status for both extant A-genome species, which, together, are equally divergent from the actual A-genome parent of the allopolyploids (Wendel et al. 1989; Wendel and Albert 1992; Cronn et al. 1996; Liu et al. 2001b). Initial evidence based on restriction site data and Southern hybridization analysis of cytoplasmic DNA (Wendel 1989; Galau and Wilkins 1989; Small and Wendel 1999) identified the A-genome as the cytoplasmic donor, i.e., maternal

parent. This was later confirmed by numerous analyses of complete chloroplast (Xu et al. 2012; Chen et al. 2017b) and mitochondrial (Tang et al. 2015; Chen et al. 2017c) genomes. An additional unexpected consequence of the numerous phylogenetic analyses devoted to understanding the origin of the polyploid species is the realization that the extant A-genome species better represent the actual A-genome parent than *G. raimondii* which represents the D-genome parent, by a factor of two.

Following their initial origin, allopolyploid cottons rapidly radiated into three major lineages (Wendel 1989; Percy and Wendel 1990; DeJoode and Wendel 1992; Stanton et al. 1994; Wendel et al. 1995a; Cronn et al. 1996; Small and Wendel 1999; Grover et al. 2012a, 2015a; Gallagher et al. 2017) while spreading throughout the coastal tropical and subtropical regions in the Caribbean, northern South America, and Central America, with some very distant extensions into the Pacific Ocean (Brubaker and Wendel 1994; Wendel et al. 2010). At present, the allopolyploid AD genome cottons contain seven recognized species (Wendel and Grover 2015; Wang et al. 2018a): $(AD)_1$ *G. hirsutum*; $(AD)_2$ *G. barbadense*; $(AD)_3$ *G. tomentosum*; $(AD)_4$ *G. mustelinum*; $(AD)_5$ *G. darwinii*; $(AD)_6$ *G. ekmanianum*; and $(AD)_7$ *G. stephensii.* The latter two comprise recently described additions that were both newly collected and cryptically archived in germplasm collections as variants of *G. hirsutum* (Krapovickas et al. 2008; Grover et al. 2015c; Gallagher et al. 2017).

Relationships among the polyploid species are relatively well understood, despite the challenges inherent in characterizing rapid radiations. The earliest divergence (~0.63 mya estimated by Chen et al. 2020) separates the lineage of *G. mustelinum* from the rest of the polyploids, which subsequently split into the *G. barbadense-G. darwinii* clade and the *G. hirsutum-G. tomentosum* complex, the latter of which also includes the two newly assigned species *G. ekmanianum* and *G. stephensii*. It is ecologically noteworthy that most of the species are island endemics (with the exception of *G. hirsutum* and *G. barbadense*) which originated following long-distance dispersal events. The two newly described species *G. ekmanianum* and *G. stephensii*, which are sister species to *G. hirsutum*, are island endemics from Hispaniola (Grover et al. 2015a) and Wake Atoll (Gallagher et al. 2017), respectively. *Gossypium darwinii* is native to the Galápagos Islands, where it may form large and continuous populations in some areas (Percy and Wendel 1990). *Gossypium tomentosum*, on the other hand, is native to the Hawaiian Islands, where it has a more diffuse population structure, occurring mostly as scattered individuals and small populations on several islands (DeJoode and Wendel 1992). Even the earliest diverging allopolyploid, *G. mustelinum*, has an island-like distribution in the sense that it is an uncommon species restricted to a relatively small region of northeast Brazil (Wendel et al. 1994). Notably, it is only the two cultivated species, *G. hirsutum* and *G. barbadense*, that have large indigenous ranges. *Gossypium hirsutum* is widely distributed in Central and South America and the Caribbean and even reaches distant islands in the Pacific (Solomon Islands, Marquesas), whereas *G. barbadense* has a more southerly indigenous range, centered in the northern third of South America but with a large range of region overlap with *G. hirsutum* in the Caribbean. These species also encompass a wealth of morphological forms that span the wild-to-domesticated continuum (Fryxell 1979; Wendel et al. 1992; Brubaker and Wendel 1994;

Brubaker et al. 1999a). Both species possess truly wild forms, although limited in *G. barbadense*, which are important in research and to our understanding the process of domestication in these two species.

2.5 Polyploidy and Evolutionary Genomics of Cotton

Polyploidy is a phenomenon that is common among plants and which can fundamentally alter the evolutionary trajectory of the incipient species. The biological, ecological, and evolutionary consequences of polyploidy are potentially numerous and lineage dependent (Madlung 2013; Soltis and Soltis 2016), and the underlying molecular changes can be remarkably complicated. These molecular changes include those that immediately operate upon genome doubling (or shortly thereafter), as well as those that operate to return polyploid genome to a modified, diploid-like state. The realization that all flowering plants are paleopolyploid (Jiao et al. 2011) indicates that these processes are repeated and cyclical (Wendel 2015).

As early as the 1930s, not only were allopolyploid cotton species identified as such, but diploid cotton itself was suggested as paleopolyploid based on the observation of secondary associations during meiotic metaphase (Lawrence 1931; Davie and Hugh Davie 1933; Skovsted 1933, 1937). Classical cytogenetic data and molecular experiments led to the speculation that $n = 7$ was ancestral to the tribe, possibly even to the Malvaceae (Davie and Hugh Davie 1933; Abraham et al. 1940; Saunders 1961; Seelanan et al. 1999; Muravenko et al. 1998). This supposition suggested a minimum of one polyploidization for diploid cotton ($n = 13$). This observation remained unconfirmed until the publication of the *Gossypium raimondii* genome (Paterson et al. 2012), which revealed a surprisingly complex structure of five- to sixfold ploidy increase approximately 60 mya after the divergence of the cotton family and its many allies (Malvadendrina, including the traditionally recognized Malvaceae (Conover et al. 2019) from chocolate (*Theobroma cacao*, Malvaceae subfamily Byttnerioideae). Subsequent analyses indicate that this multiple polyploidy history resulted not from a single event but through multiple successive events (Wang et al. 2016b), although pinpointing the lineages involved in the paleopolyploidy of cotton remains challenging due to the complicated evolutionary history of genome doubling in the Malvaceae (Conover et al. 2019). Extensive gene loss (c. 70%) following the ancient whole-genome multiplication event(s) is inferred, given the current gene count for the modern *G. raimondii* genome. It seems likely that this gene loss was nonrandom with respect to the progenitor genome contributions; that is, more gene losses were observed in the duplicated regions having lower level of gene expression and higher density of transposable elements (Renny-Byfield et al. 2015). This phenomenon of biased fractionation is in concordance with those originally identified in maize (Schnable et al. 2009, 2011; Woodhouse et al. 2010) and *Brassica* (Wang et al. 2011c; Tang et al. 2012; Cheng et al. 2012) .

For contemporary allopolyploid species, the molecular consequences of genome doubling are more readily apparent. Previous research has characterized the suite of

changes accompanying polyploidy in *Gossypium*, including gene loss and silencing, mobilization of transposable elements, epigenetic modifications, and massive genome-wide transcriptomic responses (previously reviewed by Adams et al. (2009), Wendel et al. (2012) and Wendel and Grover (2015)). While many of these changes are more subtle than has been reported for other polyploid species, as noted by Wendel et al. (2012), both the redundant nature of polyploidy and the changes that have accompanied allopolyploidization in cotton are factors in the evolution and domestication of the species. The latter is of particular interest, given that one of the parents to the polyploid does not possess spinnable fiber (D-genome), yet appears to have consequences for the domesticated fiber phenotype (Jiang et al. 1998; Rong et al. 2007; Said et al. 2015).

2.5.1 Genome-Wide Structure Variation Upon Polyploidy

In contrast to the prevalence of chromosomal rearrangements found in other model plant allopolyploids, most notably wheat (Feldman et al. 1997; Liu et al. 1998a, b; Ozkan et al. 2001; Shaked et al. 2001) and *Brassica* (Song et al. 1995; Lukens et al. 2006), structural variation in allopolyploid *Gossypium* does not appear prominent (Gerstel 1953; Menzel and Brown 1954; Brubaker et al. 1999b; Paterson et al. 2000; Liu et al. 2001a). Perhaps as a consequence of the twofold difference in genome (and, correspondingly, chromosome) size between diploid progenitors, large-scale recombination between allopolyploid subgenomes is rare (Salmon et al. 2010; Flagel et al. 2012; Grover et al. 2017b). Although the allopolyploid genome size is not quite additive and the A subgenome has slightly smaller chromosomes than the diploid A-genome species (Davie and Hugh Davie 1933; Endrizzi et al. 1985), the twofold size difference is largely preserved between subgenomes (Skovsted 1933; Endrizzi et al. 1985; Hendrix and Stewart 2005). More recently, high-quality genome sequences have become available for most of the tetraploid cottons (Zhang et al. 2015; Chen et al. 2020 in preparation; Wang et al. 2019) as well as the model diploid progenitors (Paterson et al. 2012; Du et al. 2018). These sequences were inferred with the integration of long-range scaffolding technologies (e.g., BioNano optical mapping, high-throughput chromosome conformation capture data (Hi-C), etc.), which provide a foundation for detailed comparative genomics and characterization of genome-wide structure variation. While broad-scale colinearity is largely preserved among diploid and polyploid cottons, comparisons among genome sequences have identified smaller-scale structural differences that were previously uncharacterized. For example, comparisons between the *G. hirsutum* and *G. barbadense* genomes found extensive accumulation of pericentric and paracentric inversions (Zhang et al. 2015; Yuan et al. 2015; Wang et al. 2019) in one relative to the other. With the release of additional polyploid cotton genomes, the phylogenetic placement and polarization of these events becomes possible. Such analyses will also facilitate understanding of the evolutionary and functional consequences of these smaller-scale structure variations.

2.5.2 Dynamics of Repetitive Elements

The dynamic genomic response observed in polyploid genomes is frequently associated with activities and alterations in the transposable element (TE) fraction of the nascent allopolyploid genome. Activation of TEs during hybridization and polyploidization is a long observed phenomenon (McClintock 1984; Sarilar et al. 2011; Parisod and Senerchia 2012; Piednoël et al. 2013; Senerchia et al. 2014; An et al. 2014; Vicient and Casacuberta 2017) that is linked to decreased repression and/or increased transposition (Kashkush et al. 2003; Madlung et al. 2005; Kawakami et al. 2010; Lopes et al. 2013; Ågren et al. 2016; Springer et al. 2016). Consequences for the genome include altered gene expression levels and genomic rearrangements (reviewed by Parisod and Senerchia (2012) and Vicient and Casacuberta (2017)), both of which are common to polyploid species and the latter of which may contributed to fractionation and/or diploidization (Vicient et al. 1999; Freeling and Thomas 2006; Bruggmann et al. 2006; Lim et al. 2007; Woodhouse et al. 2010, 2014; Vicient and Casacuberta 2017).

Some polyploids, such as allopolyploid cotton, exhibit relative quiescence in TEs post polyploidization (Liu et al. 2001a; Ben-David et al. 2013; Sarilar et al. 2013). As in many other plant species, genome size variation among diploid cottons is largely attributable to specific families and classes of dispersed repetitive elements (e.g., *gypsy* and *copia* LTR retrotransposons), which have differentially proliferated in different *Gossypium* lineages (Hawkins et al. 2006, 2008; Renny-Byfield et al. 2016; Grover et al. 2017a, 2019a). Following allopolyploid formation, however, there is no phylogenetic evidence to support any quantitatively significant proliferation of TEs (Hu et al. 2010). Also, whole-genome sequencing suggests that the TE composition of the diploid progenitors is mainly retained in the allopolyploids, using *G. hirsutum* as a model (Zhang et al. 2015). Despite the absence of massive bursts of transposition, evidence of post-polyploidization TE activity has been found using fluorescent in situ hybridization (Zhao et al. 1998; Hanson et al. 1998, 2000). That is, both Hanson et al. (1998) and Zhao et al. (1998) found that A-genome-specific repetitive elements have spread to the D subgenome following allopolyploid formation. More recently, de novo genome sequences of multiple allopolyploid cotton species (Chen et al. 2020 in preparation) suggest a decrease in the repetitive content in the A subgenome relative to its progenitor and, conversely, a higher repetitive fraction in the D subgenome relative to its diploid progenitor. This is consistently observed in all five sequenced allopolyploid genomes and, together with the observation that the allopolyploid genome size is slightly less than the sum of the two diploid progenitors, may suggest cryptic turnover in TE elements that includes one or more episodes of directionally biased, inter-subgenome TE transpositions and/or biased genomic removal of TE sequence in spite of an overall downsizing.

2.5.3 Fates of Duplicated Genes by Polyploidy

Polyploidy is predicted to reduce the selective pressures on individual genes compared to their orthologs in diploid progenitors, resulting in the opportunity for three classically recognized flavors of functional evolution. A gene can gain a new

function relative to its orthologs in the diploid progenitors (neofunctionalization); homoeologous gene pairs can split their ancestral function (subfunctionalization) either at the biochemical level (e.g., enzymes with two active sites for two different substrates can split each function between the two homoeologous proteins) or at the gene expression level (e.g., homoeologs are differentially expressed across different developmental times and/or tissues); or one of the two gene copies can be lost (non-functionalization). Among the earliest plant examples of subfunctionalization were those in cotton (Adams et al. 2003), where A- and D-homoeologous copies (thereafter, At and Dt) of *adhA* were shown to be reciprocally silenced in different whorls of the same flower and many other genes showed strongly biased homoeologous expression (Adams et al. 2004; Liu and Adams 2007; Flagel et al. 2008; Chaudhary et al. 2009; Adams and Wendel 2013).

Although the molecular evidence is still rather scarce regarding functional evolution of At and Dt homoeologs (e.g., Zhao et al. 2018), we have known a great deal about the various outcomes of their sequence evolution from a phylogenetic perspective. Specifically, if homoeologs evolve independently following allopolyploid formation, they should be each phylogenetically sister to the orthologous copy from their diploid progenitors, rather than to each other; falsification of this null hypothesis is indicative of interactions between homoeologs, such as gene conversion, gene loss, and unequal evolutionary rates between homoeologs and their orthologs in the progenitor diploids. This topic has been reviewed (Wendel and Cronn 2003; Adams et al. 2009; Wendel et al. 2012), but only prior to the recent explosion of genomic data; here we provide an updated synopsis.

Homoeolog sequences may evolve in a "concerted" fashion, as mediated by gene conversion or other mechanisms of homoeolog exchange. Following an early report of bidirectional inter-subgenomic homogenization of ribosomal genes (Wendel et al. 1995b), Cronn et al. (1999) studied the homoeologous sequences for 16 nuclear genes and concluded that in principle, single- or low-copy number genes have evolved independently between A and D subgenomes. In the last decade, various sources of large-scale sequence data, from expression sequence tags (Salmon et al. 2010; Flagel et al. 2012) to genomic sequencing (Guo et al. 2014; Li et al. 2015; Page et al. 2016), have been used in an effort to assess the extent of nonindependent homoeolog evolution in allopolyploid *Gossypium*. Noting the methodological caveats and analytic artifacts that could lead to overestimation of homoeolog gene conversion, it is likely that the proportion of the allopolyploid genome that has experienced homoeolog conversion is low, most likely well below 5% (Page et al. 2016).

When homoeologous loci appear to evolve independently, one key question regarding genic evolution is whether evolutionary forces and rates have been equivalent for duplicated genes within the same nucleus, and whether this is true in comparison with the orthologous copies from the progenitor diploids. Using small numbers of nuclear genes, early work first demonstrated a potential enhancement of evolutionary rate in both subgenomes of allopolyploid *Gossypium* relative to its diploid progenitors (Cronn et al. 1999; Senchina et al. 2003). This speculation was supported by subsequent molecular evolutionary analyses at much larger scales, based on transcriptome (Flagel et al. 2012) and whole-genome sequencing data (Zhang et al. 2015; Hu et al. 2019). In the latter studies, Zhang et al. (2015) and Hu et al. (2019) also found that the A subgenome appears to evolve faster than the D

subgenome, according to the distribution of synonymous and nonsynonymous substitution rates (Ks and Ka, respectively) estimated from over 20,000 homoeologous and orthologous gene sets. Interestingly, the opposite conclusions were reached by Li et al. (2015) from a separate assembly of the *G. hirsutum* genome, finding that the diploid A and D genomes are evolving faster than their orthologs in *G. hirsutum* and that the Dt subgenome is evolving faster than the At subgenome. While these contradicting results are likely rooted in the methodology used for determining evolutionary rates, comparing the evolutionary rates between multiple allopolyploid species will be useful in determining the generalities of these conclusions.

Since allotetraploid formation, the majority of At and Dt homoeologs have been retained, and this duplicate gene retention is widely thought to provide raw materials for the origin of the morphological, physiological, and ecological novelty of the allopolyploids. Despite the observation that gene loss in the allopolyploid cotton genome is rare, the *G. hirsutum* genome sequences indicate that gene loss is asymmetric between A and D subgenomes; that is, more genes have been lost in the A subgenome than in the D subgenome (228 vs 141 losses in Zhang et al. (2015) and 523 vs 461 losses in Li et al. (2015), respectively). These observations suggest that gene loss is most likely an ongoing, biased process in allopolyploid cotton. More recently, based on de novo genome sequences of multiple allopolyploid cotton species (Chen et al. 2020 in preparation), the net At gene loss is threefold the Dt loss prior to the diversification of allopolyploids, suggesting that biased gene loss is immediate following the formation of allopolyploid cotton. Whether this biased loss is stochastic or destined by "genomic legacy" of the parental diploid genomes is a matter for further investigation. Similarly, the relevance of these patterns of gene loss and retention is of particular interest to advance our understanding of phenotypic novelty accompanying allopolyploidy and domestication.

2.5.4 Evolution of Duplicate Gene Expression

In addition to sequence evolution, homoeolog genes can exhibit divergent expression patterns, a phenomenon that is widely recognized as "homoeolog expression bias" in allopolyploids (Grover et al. 2012a; Yoo et al. 2014). Following the initial report of this phenomenon (Adams et al. 2003), many aspects of this observation have been examined in allopolyploid *Gossypium* (Adams et al. 2009, 2012; Adams and Wendel 2013), including the scope and scale of biased homoeolog expression, its context specificity (e.g., tissue-specific and stress-related), the genome-wide bias, the associated phenomenon of expression-level dominance (Rapp et al. 2009), revamping of coexpression networks (Gallagher et al. 2016; Hu et al. 2016), and the temporal scale at which alterations in homoeolog expression evolve. These studies demonstrate that homoeolog bias is the rule rather than the exception in cotton, and likely most other allopolyploids. In spite of the sweeping nature of this process, virtually nothing is known about its physiological and evolutionary relevance, although clues are beginning to emerge.

At the transcriptome level, 20–55% of homoeologous gene pairs exhibit A- or D-biased expression across multiple tissues and developmental conditions (Yoo and Wendel 2014; Hovav et al. 2015; Zhang et al. 2015). This dynamic range is in concordance with early findings on the tissue and organ specificity of homoeolog expression changes (Adams et al. 2004; Liu and Adams 2007; Chaudhary et al. 2009). Regarding the genome-wide pattern, neither A or D subgenome appears globally dominant in terms of biased homoeolog expression in all tissues. That is, the numbers of A- and D-biased genes often are more or less balanced, as during fiber development (Yoo and Wendel 2014). For example, more A-biased genes were found in ovules at the stage of fiber initiation (i.e., 3 days pre- and post-anthesis) (Yang et al. 2006), but the reverse was observed in seeds (Hovav et al. 2008). In addition, both balanced pattern (Rambani et al. 2014) and unbalanced bias favoring the D subgenome (Flagel et al. 2008) have reported for the petal tissue. The most sweeping study to date, in this respect, was by Zhang et al. (2015), who demonstrated a preferential utilization of the D subgenome in 31 of the 35 tissue/stage samples in *G. hirsutum.* Although some have speculated that genome-wide bias may have been selected to facilitate tissue-specific function and morphology that is more similar to progenitor of the favored subgenome, this doesn't seem to be the case in *Gossypium* given the fact that D-genome diploid does not produce spinnable fibers.

Whereas the mechanistic basis of homoeolog expression bias remains to be elucidated, experimental work does reveal that it can be attributed to two temporal stages, the changes that ensue immediately upon hybridization and those arising later from polyploidy and subsequent evolution (Flagel et al. 2008, 2012; Yoo et al. 2013; Rambani et al. 2014). One consensus observation from these studies, as well as in other allopolyploid plants (Xu et al. 2014; Wang et al. 2016c; Sun et al. 2017), is that a considerable proportion of the expression biases that accompany hybridization become maintained during allopolyploid evolution, and the initial bias often becomes amplified on an evolutionary timescale. Differential expression of homoeologs accompanying hybridization of diverged diploid genomes is diagnostic of *cis*-regulatory divergence between the parental genes, because the common *trans* effect alone should lead to equal homoeolog expression in the absence of *cis* differences (Wittkopp et al. 2004). Hu and Wendel (2019) recently pointed out the opportunities for understanding allopolyploid gene expression and hence phenotypes, through extending the classic *cis-trans* model to explicitly incorporate genome doubling.

Distinct from homoeolog-specific expression, a series of other expression patterns have also been studied for the aggregated expression of a given homoeologous pair, such as additive and nonadditive expression, expression-level dominance, and transgressive expression, as previously reviewed (Grover et al. 2012a; Yoo et al. 2014; Hu and Wendel 2019). About 23–61% of homoeologous pairs exhibited nonadditive expression among variable cotton tissues (Flagel and Wendel 2010; Yoo et al. 2013; Rambani et al. 2014). The phenomenon of expression-level dominance was originally discovered and elaborated in cotton (Rapp et al. 2009; Flagel and Wendel 2010), defined as the state where total expression of a homoeolog pair mimics the expression level of one of the two diploid parents of an allopolyploid; this parent is referred to as "dominant." At the genome-wide scale, Rapp et al. (2009)

demonstrated that in the synthetic AD allopolyploid ($2(A_2D_1)$), more than 10,000 genes are D-dominant, significantly higher than ~800 A-dominant genes. At present, the functional significance is unknown, but it's scale and scope suggest that it is important from many perspectives, including crop plant improvement. Both homoeolog expression bias and expression-level dominance also have been reported at the protein level (Hu et al. 2011, 2013, 2015).

Beyond the gene-centric characterization of expression changes, coexpression network analysis in polyploids has the potential to facilitate a better understanding of the mechanistic underpinnings of phenotypic and ecological traits, and also may provide novel insight into interactions among duplicated genes and genomes (Gallagher et al. 2016). Hu et al. (2016) conducted coexpression network analysis of developing seeds from diploid and allopolyploid cotton species and found that the network topology in polyploids is to some extent a modular combination of that of its progenitor genomes. Interestingly, the allopolyploid network resembles the network of the A-genome diploid more than that of the D-genome parent, despite its D-like phenotype in oil content. Expression modifications for entire modules of gene expression include analogies to phenomena described for single pairs of duplicated genes, i.e., coexpression-level dominance and transgressive expression.

2.5.5 Epigenetic Modifications

Epigenetic modifications are another consequence of polyploidization, particularly for allopolyploids, whose formation includes hybridization of divergent species. These modifications include DNA methylation, histone modifications, and chromatin remodeling, among others, whose patterns of change under polyploidization are reviewed elsewhere (Chen 2007; Paun et al. 2007; Song and Chen 2015; Vicient and Casacuberta 2017; Ding and Chen 2018).

While the epigenetic consequences of allopolyploidization in cotton are understudied, results from methylation surveys suggest that the allopolyploid cotton genome is epigenetically stable (Liu et al. 2001a). This observation was reiterated in later surveys, which also found that homoeologous methylation changes correlated with expression bias (Song et al. 2015, 2017). Furthermore, homoeologous demethylation of the *COL2D* gene in cultivated cottons was associated with photoperiod insensitivity (Song et al. 2017), demonstrating the potential of evolutionarily relevant epigenetic modifications in allopolyploid cotton. The relevance of methylation patterns to agronomic traits in allopolyploid cotton has been considered for nearly 50 years, when demethylation was associated with *Verticillium* stress in *G. hirsutum* (Guseinov et al. 1975). Subsequent surveys have noted that demethylation is also associated with salt tolerance (Zhao et al. 2010; Wang et al. 2016a) and the response to cold stress (Fan et al. 2013) and that methylation patterns can influence fiber growth (Jin et al. 2013). Given that methylation diversity in *G. hirsutum* is higher than genetic diversity (Keyte et al. 2006; Osabe et al. 2014), and that methylation can be significantly different among tissues (Osabe et al. 2014), further

understanding of the patterns of methylation among species, accessions, and tissues can provide insight into biased gene expression patterns (Adams et al. 2003; Flagel and Wendel 2010; Grover et al. 2012a; Song et al. 2017) and, ultimately, phenotypes. Although other forms of epigenetic modification are far less understood for cotton, initial research into histone modifications (Zheng et al. 2016), microRNAs (Guan et al. 2014), and long-noncoding RNAs (Wang et al. 2015a) suggests that understanding these may also provide insight into how the homoeologous genomes of allopolyploid cotton are used and/or interact, with consequences for fiber (Guan et al. 2014) and other phenotypic traits.

2.6 Genomes of the Four Domesticated Species

2.6.1 Diversity and Origin

A remarkable feature of the cotton genus is that domestication has occurred not once but independently four times, involving two allopolyploid species from the Americas, *G. hirsutum* and *G. barbadense*, and two diploids from Africa-Asia, *G. arboreum* and *G. herbaceum*. In each of these four cases, aboriginal people discovered several thousand years ago the unique properties of cotton fibers and made them useful for ropes, textiles, and other applications. As a consequence, all four species are currently grown worldwide, occupying 30–36 million hectares globally in 30 countries (Kranthi 2018) and with smaller shares in more than 70 additional countries (Chee et al. 2016). Cotton fiber production comprises a multibillion dollar industry responsible for hundreds of millions of jobs annually. Accordingly, the history and impact of domestication on these four species is of substantial interest.

2.6.2 Domesticated Diploids

As noted above, both extant species of the A-genome (i.e., *G. herbaceum* and *G. arboreum*) have been domesticated for fiber production. Collectively known as the "short staple cottons" (Khadi et al. 2010), both diploid species produce comparably short and coarse fibers that are primarily used for handloom textiles, hosiery, fillers (e.g., mattress stuffing), and absorbent materials (Kranthi 2018). Although these are the least grown cotton species, they are of regional importance due to pest and/or environmental resistance (Basu 1996; Rajendran et al. 2005). Commercial production of these species is ongoing in India, Pakistan, Iran, Myanmar, and Thailand; *G. arboreum* alone is 22% of the cotton in Myanmar and over 50% in Thailand, and *G. herbaceum* occupies up to 0.6 million hectares in India (Kranthi 2018). Because the importance of these species as cultivars is limited to a few countries, our understanding of the extant diversity is generally limited.

Of the two diploids, only *G. herbaceum* has a known wild form, native to the savannahs of Southern Africa (Vollesen 1987; Wendel et al. 1989; Khadi et al. 2010). While the center of domestication of *G. herbaceum* is unclear, it likely was in Northern Africa or the Near East as pointed out by Fryxell (1979). Range expansions extended the initial growing range through the Persian Gulf states and the Indian subcontinent (Kranthi 2018). Genetic diversity in *G. herbaceum* relative to its sister species is unclear, with some reporting lower diversity in *G. herbaceum* (Wendel et al. 1989), whereas others report higher diversity (Jena et al. 2012). These conflicting observations are attributable to differences in germplasm evaluated, as well as markers (i.e., allozymes versus AFLP markers). Morphological diversity has also been characterized (Wendel et al. 1989), which, together with geographical factors, has partitioned the species into five infraspecific races (Stanton et al. 1994; Khadi et al. 2010); however, the utility of these races is unclear, and the partitioning of new accessions into a particular race is highly subjective.

Diversity in *G. arboreum* is similarly understudied, although the few estimates also indicate that diversity is low (Page et al. 2013; Fang et al. 2017a). Colloquially referred to as "Desi cotton" in the Indian subcontinent, *G. arboreum* has no true wild forms and was once considered a possible derivative of *G. herbaceum* that became separated as the result of a reciprocal translocation (Gerstel 1953). While recent research suggest that *G. arboreum* underwent speciation from its sister species, *G. herbaceum*, long prior to domestication (Renny-Byfield et al. 2016; Du et al. 2018), there is little known about the wild ancestors of this cultivated species and/or its domestication and diversification into the modern cultivated forms. The Indus Valley has historically been suggested as the origin for the species, as it represents a center of diversity and has archaeological evidence (Gulati and Turner 1928; Wendel et al. 2010). This, however, may actually represent a secondary center of diffusion post-domestication (Hutchinson 1954; Wendel et al. 2010). More recent analyses of resequencing data place the origin of Chinese accessions in South China with subsequent radiation to the Yangtze and Yellow River regions (Du et al. 2018). Traditionally, *G. arboreum* production has been limited to Asia, ranging from India to Korea (Wendel et al. 1989; Basu 1996; Guo et al. 2006; Khadi et al. 2010). While accounting for only 1–3% of the total cotton-growing area in India, *G. arboreum* occupies far more of the growing area in Myanmar and Thailand (Kranthi 2018). As with *G. herbaceum*, *G. arboreum* has been traditionally partitioned into five botanical races, but these appear to have limited usefulness. As mentioned above, relative diversity within *G. arboreum* is largely unclear, including diversity compared to tetraploid cultivars (see below). That is, while there exists some evidence that the amount of diversity in the diploid species is roughly equivalent to that found in the tetraploids (Wendel et al. 1989; Stanton et al. 1994), others suggest that the diploid cultivars retain more diversity than do the polyploid species (Jena et al. 2012), possibly due to more intense selection in the latter. Clearly, additional research on diversity within and among these two species is warranted, to facilitate an understanding of their modern gene pools, patterns of interspecific introgression (Wendel et al. 1989), and their domestication bottlenecks.

2.6.3 Domesticated Polyploids

Most famously, the cotton genus has also given us two domesticated polyploid species, *G. hirsutum* and *G. barbadense*, collectively responsible for the vast majority of the cotton trade. As noted above, both species are derived from a single allopolyploidization event arising from the chance dispersal of an A-genome ancestor to the Americas, one that fortuitously bore long fibers on its seeds. Interestingly, while the contribution of the African A-genome is apparent in both wild and domesticated fiber phenotypes of all seven wild allopolyploid species, the contribution of the short-fibered paternal parent to the allopolyploid fiber is more enigmatic, although potentially important (Reinisch et al. 1994; Jiang et al. 1998, 2000; Lacape et al. 2005; Han et al. 2006; Rong et al. 2007; Said et al. 2015). Although independently domesticated, the naturally occurring range of *G. hirsutum* and *G. barbadense* overlaps in parts of Central and South America (Brubaker et al. 1999a; Wendel and Cronn 2003; Chee et al. 2016), leading to natural introgression in addition to the intentional post-domestication introgression (Wang et al. 1995; Wayne Smith and Tom Cothren 1999; May 2001; Zhang et al. 2005, 2014; Reddy et al. 2017; Fang et al. 2017a). Introgression between the two cultivars is not straightforward, as there exist partial reproductive barriers (Stephens 1946), segregation distortion (Jiang et al. 2000), and eventual hybrid breakdown (Stephens 1946; Brown and Ware 1958; Zhang et al. 2014). Nevertheless, interest remains in combining the most desirable traits from each species (Zhang et al. 2014; Cao et al. 2015).

G. hirsutum or "upland cotton" is the most widely cultivated cotton species, accounting for 98–99% of the total cotton-growing area (Kranthi 2018). Native to Mexico and Central America, the initial domestication center of *G. hirsutum* is unclear, although northern Mesoamerica, specifically the Yucatán Peninsula, *circa* 5000 years ago, has been proposed (Wendel and Albert 1992; Wendel et al. 1992; Brubaker and Wendel 1994; Chee et al. 2016). The history of transformation from dooryard cultigen into a modern row crop is better documented, having occurred in the southern United States of America in what is known as the "Cotton Belt" (Ware 1951). Domestication of *G. hirsutum* involved the transformation of wild forms into those that bore commonly domesticated traits, e.g., compact plant architecture, day-length neutrality, morphological exaggeration of desirable phenotype (here, fiber), etc. Naturally, the intense selective pressure during domestication resulted in a reduction in genetic diversity; however, the reduction in diversity for domesticated cotton exceeds what is commonly observed for domesticated species (Wendel et al. 1992; Brubaker and Wendel 1994). The low levels of diversity in *G. hirsutum* have been characterized using various markers (Hutchinson 1951, 1959; Wendel 1989; Wendel et al. 1992; Brubaker and Wendel 1994, 2001; Iqbal et al. 1997, 2001; Brubaker et al. 1999a; Lubbers and Chee 2009; Ahmad et al. 2012; Fang et al. 2013; Tyagi et al. 2014; Zhao et al. 2015; Grover et al. 2017b), which all support a very narrow germplasm base.

While *G. barbadense* represents a significantly smaller share of cotton grown, it is frequently desired for its longer, stronger, and finer fiber (Liu et al. 2015b).

Colloquially known as "Egyptian" or "Pima" cotton, *G. barbadense* was domesticated in the dry coastal regions of northern Peru and southern Ecuador at least 7500 years ago (Percy and Wendel 1990; Westengen et al. 2005; Splitstoser et al. 2016). Following a primary domestication west of the Andes, the range of *G. barbadense* was expanded into northern South America through a trans-Andean dispersal and then into Central America, the Caribbean, and the Pacific (Percy and Wendel 1990; Rossen et al. 1996; Piperno and Pearsall 1998; Westengen et al. 2005). Modern cultivars of *G. barbadense* were first developed on the coastal plains and islands of the southeastern USA (as "sea island cotton") and were subsequently introduced to Egypt, where *G. barbadense* acquired the designation "Egyptian cotton" (Khadi et al. 2010). Subsequent reintroduction of *G. barbadense* (as "pima cotton") into the southwestern USA gave rise to the modern elite cultivars, also known as "extra-long-staple" cotton (Khadi et al. 2010). While the overall market share of *G. barbadense* is less than that of *G. hirsutum*, some countries (e.g., Egypt and Israel) favor the superior fiber quality, and plant this species mostly or entirely exclusively (Kranthi 2018). Like *G. hirsutum*, multiple accessions that span the wild to domesticate continuum are available, including early dooryard cultigens and landraces (Percy and Wendel 1990); however, truly wild forms of *G. barbadense* are rare. Diversity in this species has been characterized (Percy and Wendel 1990; Westengen et al. 2005; Lacape et al. 2007; Boopathi et al. 2008; Wang et al. 2011b; Abdellatif et al. 2012; Hinze et al. 2015, 2016; Grover et al. 2017b), although the number of modern datasets is low (Hinze et al. 2016). In general, it is accepted that *G. barbadense* also has low diversity that is either similarly low (Page et al. 2016; Grover et al. 2017b) or lower than (Abdalla et al. 2001; Lacape et al. 2007; Hinze et al. 2015, 2016) that found in *G. hirsutum* (however see Van Deynze et al. 2009).

2.7 Whole-Genome Resequencing Studies and Insights into the Genetic Structure of the Domesticated Species

Advances in next-generation sequencing (NGS) and third-generation sequencing technologies have greatly facilitated the sequencing of the domesticated cotton genomes (Wang et al. 2012, 2019; Paterson et al. 2012; Li et al. 2014, 2015; Zhang et al. 2015; Liu et al. 2015b; Yuan et al. 2015; Du et al. 2018). These studies collectively provide an important, enabling resource for a wealth of opportunities in plant breeding, genomic research, and gene discovery, among other applications. Following the initial completion of reference genome sequences (reviewed by Wang et al. 2015b), several whole-genome resequencing (WGR) studies have been reported (Page et al. 2013, 2016; Fang et al. 2017a, b; Wang et al. 2017; Du et al. 2018; Ma et al. 2018), which provide unprecedented insights into the evolutionary relationships between different gene pools, the structuring of genome variation, and revealing variable types of diversity (such as single-nucleotide polymorphism and InDels) within populations. Here, we summarize the recent progress in cotton WGR

studies with a special emphasis on their applications to understand the genetic basis of phenotypic variation (Table 2.2).

The first studies of whole-genome resequencing in diploid and tetraploid cotton were reported by Page et al. (2013, 2016). A total of 10 accessions of diploid cotton and 34 accessions of tetraploid cotton were resequenced separately, with an average coverage of about 37× and 23× per accession, respectively. In 2017, WGR efforts in cotton accelerated, with three massive datasets being released. Fang et al. (2017a) sequenced 147 accessions of cotton (a total of 1.8 terabases) with an average of approximately 5× coverage per accession, including 33 accessions of *G. hirsutum* landraces, 52 *G. hirsutum*, 52 *G. barbadense* cultivars, and 10 other tetraploids (NCBI Project ID: PRJNA257154). Fang et al. (2017b) next sequenced 258 *G. hirsutum* cultivars (3.96 Tb total) at about 2.5× coverage each (PRJNA375965). Wang et al. (2017) sequenced 321 *G. hirsutum* cultivars (6.1 Tb total) at an average of 6.9-fold coverage (PRJNA336461). Until 2018, a total of 419 accessions of core upland cotton were sequenced which produced 6.55 Tb of data with 6.55× coverage (PRJNA399050). Complementing these efforts in tetraploid cotton, Du et al. (2018) resequenced 243 *G. arboreum* and *G. herbaceum* accessions at about a 6× coverage (PRJNA349094).

These WGR studies have permitted unprecedented insights into the patterns of variation in the cultivated cotton genomes, such as SNPs, InDels, copy number variations (CNV), and presence/absence variations (PAV), and the overall structuring of genetic diversity within and between different germplasm groups. For example, Fang et al. (2017a) found 16,377,749 SNPs in the interspecific populations of *G. hirsutum* and *G. barbadense*, among which 7,993,856 are common SNPs with allele frequency larger than 5% and a missing data rate less than 10%. The authors also discovered 144,662 InDels, which ranged in length between 1 bp and 8 kbp, and also reported 16,879 structural variations longer than 50 bp. Wang et al. (2017) detected 7,497,568 SNPs in their study of 321 *G. hirsutum* cultivars, as well as 351,013 small InDels (1–10 bps) and 93,783 structural variations (longer than 10 bps). Similar results were obtained in other WGR studies of *G. hirsutum*; for example, using 318 *G. hirsutum* cultivars, 8,621,073 SNPs were reported (Fang et al. 2017b), whereas 3,665,030 SNPs were detected for 419 core accessions of *G. hirsutum* cultivars (Ma et al. 2018). From two A-genome diploid cottons, Du et al. (2018) discovered 17,883,108 SNPs and 2,470,515 InDels.

The large sample size and sufficient sequence coverage of these WGR data allow a robust estimation of nucleotide diversity within and between populations. In the germplasm studied by Fang et al. (2017a), nucleotide diversity was reported as 0.00216 for *G. hirsutum* landraces and ~0.0007 for cultivars, indicating a strong genetic bottleneck during upland cotton domestication. Similar estimates of *G. hirsutum* cultivars were reported by the other two studies (Wang et al. 2017; Fang et al. 2017b), which consistently suggest that cultivated gene pools contain appreciable but low genetic diversity. By comparing genetic diversity and divergence between landraces and cultivars, over 200 selective sweeps—109 by Fang et al. (2017a), 93 by Wang et al. (2017), and 25 by Fang et al. (2017b)—were identified as the potential targets of human selection during crop domestication and later improvement.

Table 2.2 List of whole-genome resequencing projects

	Page et al. (2013, 2016) *G3*	Page et al. (2016) *PLOS Genetics*	Fang et al. (2017a) *Genome Biology*	Wang et al. (2017) *Nature Genetics*	Fang et al. (2017b) *Nature Genetics*	Du et al. (2018) *Nature Genetics*	Ma et al. (2018) *Nature Genetics*
Number of accessions	10	34	147	321[a]	258[b]	243	419
Species	*G. herbaceum*, *G. arboreum*, *G. raimondii*	Seven species of tetraploid cotton	*G. hirsutum* *G. barbadense*	*G. hirsutum*	*G. hirsutum*	*G. arboreum* *G. herbaceum*	*G. hirsutum*
Germplasm type	Wild, landrace, and cultivar	Wild, landrace, and cultivar	Landrace Cultivar	Cultivar	Cultivar		Cultivar
Total bases	~500 Gb	902 Gb	1.8 Tb	6.1 Tb	3.96 Tb	2.29 Tb	6.35 Tb
Sequencing depth	~37	23	~5	6.9	~2.5	~6.0	6.55
NCBI SRA ID	PRJNA202235, PRJNA202236, and PRJNA202239	PRJNA280597	PRJNA257154	PRJNA336461	PRJNA375965	PRJNA349094	PRJNA399050
Nucleotide variations (SNPs)	23,859,893 homoeo-SNPs	19.2 and 28.5 million homoeo-SNPs (35.6 million total unique loci)	16,377,749 (at least 2 accessions), 7,993,856 (MAF 0.05, missing rate 10%)	7,497,568 (depth at least 8, MAF 1%), 2,020,834 (MAF 5%)	8,621,073 2,167,186 (MAF 0.05)	17,883,108	3,665,030 (MAF 0.05, missing rate 20%), 1,980,539 (missing 10%)
InDels	–	–	144,662	351,013	–	2,470,515	–
Structural variation (SV)	–	–	16,879	93,783	–	–	–
Related traits and discovered loci	–	–	109 selective sweeps, 384 introgression events	93 domestication sweeps, 19 GWAS loci for fiber-quality-related traits, asymmetric subgenome domestication and cis-regulatory	25 improvement-selective sweeps, 119 GWAS loci (71 for yield-related traits, 45 for fiber qualities, and 3 for resistance to Verticillium wilt)	98 significant peak associations for 11 agronomically important traits	7383 GWAS SNPs for 13 fiber-related traits

[a]A total of 352 accessions, including 321 cultivated accessions and 31 wild accessions (public data)
[b]A total of 318 accessions, included 258 cultivars and other wild accessions (released previously)

An additional perspective of sequence evolution in allopolyploid cotton is provided by the comparisons of nucleotide diversity between the co-resident A and D subgenomes. In both *G. hirsutum* and *G. barbadense* cultivars, approximately equal levels of nucleotide diversity were reported (A-0.00075 vs D-0.00073 and A-0.00061 vs D-0.00051, respectively) by Fang et al. (2017a), which is consistent with a previous survey based on targeted sequence capture (Grover et al. 2012b). Interestingly, a significantly higher diversity of A than D subgenome (0.00072 vs 0.00056) in *G. hirsutum* was reported by Wang et al. (2017), which may reflect the sampling difference in the germplasms surveyed by different studies. Given that the external ecological and population-level influences should affect the co-resident genomes simultaneously, equal nucleotide diversity between subgenome is expected unless there has been selection favoring homoeologous genes from one subgenome over the other; this doesn't seem to be the case in cotton domestication.

An additional use of the resequencing data is to detect past episodes of interspecific gene flow between *G. hirsutum* and *G. barbadense*. It has long been known that these two species experienced massive historical introgression in colonial or pre-colonial times, particularly in the Caribbean and other areas where both species became intermingled (Wendel et al. 1992; Brubaker et al. 1993; Brubaker and Wendel 1994), as well as intentionally during crop improvement in the last couple of centuries (see below section: Natural and Artificial Introgression in the Genus). Some of this history is evident in the WGR data. Fang et al. (2017a), for example, reported 384 introgressed regions between *G. hirsutum* races and *G. barbadense* cultivars. They also found that the introgression events were significantly biased toward gene flow from *G. hirsutum* into *G. barbadense*, rather than the reverse. Whether these results will be confirmed by additional studies and using different germplasm datasets remains to be determined.

A central task of modern plant genetics and breeding is to understand the genetic basis of phenotypic variations. The abundant WGR resources listed above have allowed the construction of genetic variation maps and made possible genome-wide analysis studies (GWAS) as well as experiments designed to detect quantitative trait loci (QTLs) for agronomic traits. For example, Fang et al. (2017b) reported that 54.8% of the elite genome-wide association study (GWAS) alleles of upland cotton cultivars in China were transferred from three founders—Deltapine 15, Stoneville 2B, and Uganda Mian. A total of 119 GWAS loci—71 for yield-related traits, 45 for fiber qualities, and 3 for resistance to *Verticillium wilt*—were identified. Among those, the authors noted that more associated loci for lint yield and fiber quality were located in the A subgenome (70) than in the D subgenome (49), and there are more associated loci for lint yield than for fiber quality. In the study of Wang et al. (2017), a total of 19 association signals were detected for fiber-quality-related traits, including 8 and 11 in the A and D subgenomes, respectively. More recently, Ma et al. (2018) conducted a GWAS for 13 fiber-related traits, which identified 7383 unique SNPs that were significantly associated with these traits and located within or near 4820 genes; in contrast to Fang et al. (2017b), there are more associated loci identified for fiber quality than for yield, and more fiber genes detected in the D than

the A subgenome. For diploid cotton, Du et al. (2018) reported 98 significant peak associations for 11 agronomically important traits.

Overall, these examples offer a glimpse into the types of insights that will emerge from focused WGR efforts in cotton (Table 2.2). Using increasingly precise and focused high coverage datasets, an entire spectrum of questions that once were inconceivable will now become experimentally feasible. This include the many details regarding the parallel domestication of the four cultivated cotton species, insights into the genetic and genomic basis of key ecological and agronomic traits, and paths forward for continued cotton improvement in an ever-changing agricultural landscape.

2.8 Wild Cotton Species as Sources of Desirable Breeding Traits

As noted in an earlier section, the process of domestication and crop improvement were accompanied by bottlenecks that have substantially reduced genetic variation in cultivated cottons, thus limiting their potential for developing novel varieties with improved traits. In principle, their wild progenitors and other wild relatives have far more genetic diversity and likely preserve many valuable genetic variants and associated phenotypes that are not present in the crop gene pool. In the context of climate change, crop wild relatives also serve a reservoir of genetic diversity adapted to a wide range of environmental conditions that plant breeders are increasingly likely to need to create new varieties able to cope with new and possibly exceptional abiotic conditions (Dempewolf et al. 2017; Zhang et al. 2017a; Mammadov et al. 2018). In *Gossypium*, although many wild species are narrow endemics with likely low levels of within-species genetic diversity, their worldwide distribution collectively encompasses a full range of geographic and ecological variation, suggesting that there are ample resources to be explored for use in cotton breeding. Cotton improvement programs have exploited many diploid species in association with specific morphological traits, disease resistance, cytoplasmic male sterility, and fertility restoration, whereas genes for stress tolerance, disease resistance, and nectariless and glandless cotton have been deliberately introduced from wild and feral tetraploids. These genetic enhancements are mainly obtained through conventional breeding approaches involving interspecific hybridization and backcrossing. We have collected and summarized these breeding efforts (reviewed in Mammadov et al. 2018) and potentially valuable ecological traits (Hutchinson 1959; Fryxell 1979; Wendel et al. 2010) in Table 2.1, along with a list of the relevant wild species used.

For example, *G. bickii* and other Australian diploid species in section *Sturtia* and *Hibiscoidea* have greatly lessened deposition of gossypol (toxic to human beings and nonruminant animals) in seeds but retain higher levels in aboveground plant tissues (Fryxell 1965). This "glandless-seed and glanded-plant" trait was introgressed into upland cotton cultivars for maximizing the utilization of cotton seed oil and proteins while retaining resistance to many diseases and phytophagous pests

(Zhu et al. 2005). The African species *G. anomalum* provides a unique source for developing cultivars with finer and stronger fibers (Mehetre 2010), as well as improved resistance to diseases like bacteria blight and cotton rust (Fryxell et al. 1984; Endrizzi et al. 1985).

Introgression of pest and disease resistance to cultivated cotton is also exemplified by using *G. longicalyx* as a source of reniform nematode resistance (Robinson et al. 2007), using *G. stocksii* for cotton leaf curl virus (CLCuV), using *G. australe* (Liu et al. 2015a), *G. bickii* (Wang et al. 2004), *G. sturtianum* (McFadden et al. 2004), and *G. thurberi* (Zhao et al. 2012) to transfer resistance to Fusarium wilt and Verticillium wilt diseases. With respect to abiotic stress tolerance, the majority of wild diploid species are more or less xerophytically adapted to the drier environments of desert areas, except *G. longicalyx* and several little-known species from the northwest Kimberley region of Australia (Fryxell 1979). Specifically, the American D-genome species have evolved different strategies for drought tolerance: *G. harknessii*, *G. armourianum*, and *G. turneri* have evolved reduced leaves with thick cuticles and a double layer of palisade cells for reducing water loss; *G. aridum*, *G. lobatum*, and *G. laxum* stay dormant during the dry season to circumvent the aridity of the habitat. Perhaps less obvious as an adaptation, the aggressive and deeply penetrating root system enables *G. gossypioides* and *G. thurberi* to grow on steep rocky slopes far from any watercourses. Sharing many similar physiological and molecular responses to drought stress, salt tolerance was previously noted for *G. aridum*, *G. davidsonii*, and *G. klotzschianum*, and many stress-responsive genes have been identified by transcriptomic analyses (Fan et al. 2015; Zhang et al. 2016; Wei et al. 2017).

Because of the differences in ploidy, genetic incompatibility, growth habit, and other agronomic traits, transferring beneficial traits from wild diploid species into cultivated tetraploid cottons is always challenging. The most effective strategy has been developing interspecific hybrids through bridge crosses. For example, to facilitate the introgression of reniform nematode resistance, Bell and Robinson (2004) developed synthetic tetraploid triple-species hybrids between *G. hirsutum*, *G. longicalyx*, and *G. armourianum* (or *G. herbaceum*) as bridges, from which successful introgression was accomplished by recurrent backcrosses to *G. hirsutum* (Robinson et al. 2007). Such conventional breeding programs are usually time-intensive and resource-intensive. With recent advances in CRISPR/Cas genome editing in cotton (Chen et al. 2017a; Gao et al. 2017; Zhang et al. 2018; Long et al. 2018; Li and Zhang 2019; Li et al. 2019a, b), transferring specific genes from wild cotton species with improved efficiency and precision might soon become universal. Understanding the evolutionary relationship from primitive species to crops and among wild relatives will continue to provide insights into the biology of cotton and agronomic improvement.

2.9 Natural and Artificial Introgression in the Genus

The extent of natural interspecific introgression in *Gossypium* is remarkable given the number of pre- and post-reproductive barriers that exist among species (Cronn and Wendel 2003). Although the natural range of cotton species collectively

encompass much of the drier tropics and subtropics worldwide, most species exist in small, scattered populations that are geographically isolated from other regional species, making interspecific contact a rare occurrence (Cronn and Wendel 2003). For those species that are not geographically isolated, further reproductive barriers exist. Reduced pollen germination and abnormal pollen tube growth are present in some interspecific crosses (Ram et al. 2008), cytogenetic incompatibilities often lead to univalent formation, and there exist several sources of cytoplasmic male sterility (e.g., *G. harknessii* and *G. sturtianum*) that likewise inhibit interspecific crosses (Meyer 1975; Stewart 1992; Meshram 1994; Zhang and Stewart 1999, 2001; Liu et al. 2003; Suzuki et al. 2013; Chen et al. 2017d). Successful interspecific crosses, however, may also become a dead end; i.e., F1 hybrids between genome groups tend to be sterile (Endrizzi et al. 1985).

These natural barriers to introgression notwithstanding, evidence suggests that natural introgression is prevalent, although often cryptic (Cronn and Wendel 2003). Cytoplasmic introgression (i.e., introgression involving the chloroplast) appears to be most prominent. Both *G. cunninghamii* (Australian K-genome) and *G. bickii* (Australian G-genome) possess chloroplast genomes most similar to those species found in the Australian C-genome (e.g., *G. robinsonii*) (Wendel et al. 1991; Cronn and Wendel 2003), and the entire African B-genome has Australian-like chloroplasts, despite being nestled within a monophyletic African clade (i.e., A-, B-, E-, F-genomes) according to nuclear markers (Cronn et al. 2002; Cronn and Wendel 2003). In these examples, nuclear introgression has not been reported, possibly due to rapid elimination of one parental genome, which has been observed in interspecific cotton crosses (Stephens 1949, 1950).

Alternatively, nuclear introgression may be cryptically present. The first example of cryptic nuclear introgression in cotton was reported for *G. gossypioides*, involving two independent introgression events. The more recent introgression involved transfer of the *G. raimondii* cytoplasm into *G. gossypioides*, resulting in the placement of *G. gossypioides* at a terminal phylogenetic position within the D-genome chloroplast phylogeny (Small and Wendel 2000; Cronn et al. 2003; Cronn and Wendel 2003; Grover et al. 2019a). Nuclear ribosomal DNA sequences, however, strongly place *G. gossypioides* as a member of the African clade (A-, B-, E-, F-genome (Wendel et al. 1995a; Seelanan et al. 1999; Cronn et al. 2003)), which is supported by the presence of African-specific repetitive sequences in *G. gossypioides* (Zhao et al. 1998). Low-copy nuclear sequences, however, suggest that *G. gossypioides* is the first lineage to diverge within the D-genome (Small and Wendel 2000; Liu et al. 2001b; Cronn et al. 2003; Grover et al. 2019a). Together, these suggest a more ancient introgression of an African cotton into *G. gossypioides* that includes nuclear introgression, followed by a more recent cytoplasmic introgression of the *G. raimondii* chloroplast into *G. gossypioides*.

Similarly, populations of *G. aridum sect. Erioxylum* from the Mexican state of Colima have experienced cryptic introgression from section *Houzingenia.* This population of *G. aridum* is distinct from other populations (e.g., those from Jalisco, Mexico) in that the chloroplast is more similar to the sister species *G. klotzschianum* and *G. davidsonii* than it is to the remaining members of section *Erioxylum* (Alvarez

and Wendel 2006; Grover et al. 2019a). Recently, genome-wide analyses of *G. aridum* with the remainder of the D-genome species also detected remnants of nuclear introgression (Grover et al. 2019a), which was not previously reported. Other instances of cryptic nuclear introgression in diploid *Gossypium* may become apparent as resequencing within the genus becomes commonplace. Candidates include *G. triphyllum*, one of the B-genome cottons with Australian cytoplasm that also exhibits some morphological similarities to the Australian cottons (Fryxell 1979; Cronn and Wendel 2003), and *G. bickii*, which as noted above appears to have a hybrid ancestry.

Naturally occurring introgression among polyploid species also has been reported, although this seems unlikely for more restricted island endemics like *G. tomentosum* (Hawaii) and *G. darwinii* (Galapagos Islands). Perhaps the best characterized is the introgression between *G. hirsutum* and *G. barbadense* in the Caribbean and northern South America (Stephens 1967; Percy and Wendel 1990; Wendel et al. 1992; Brubaker et al. 1993; Brubaker and Wendel 1994; Coppens d'Eeckenbrugge and Lacape 2014; Hinze et al. 2016). Despite the reproductive barriers mentioned above, bidirectional introgression has been detected for populations of both species in their overlapping ranges. The amount of introgression, however, is not equivalent between the two species; that is, *G. hirsutum* naturally carries more *G. barbadense* alleles than the converse (Chee et al. 2016). This may be due to the aforementioned phenomenon of segregation distortion (Jiang et al. 2000) and/or biased genome elimination (Stephens 1949, 1950). This ability of *G. hirsutum* to integrate *G. barbadense* alleles may have led to the creation of the botanical race Marie-Galante (Stephens 1967, 1974), and further introgression of the Brazilian endemic *G. mustelinum* into some populations of Marie-Galante may have subsequently resulted in the Brazilian "moco" cotton (Pinto de Menezes et al. 2010).

In addition to natural interspecific gene flow, humans have introduced intention introgression into breeding stocks for cultivar development (Saha et al. 2006; Wang et al. 2011a; Zhang et al. 2014; Chee et al. 2016). As mentioned above, some are from diploid sources. In particular, there is much interest in using introgression between *G. hirsutum* and *G. barbadense* to enhance fiber traits. To date, it appears that there has been more success introgressing *G. hirsutum* into *G. barbadense* than the converse (Chee et al. 2016), converse to the natural direction.

2.10 Cotton Genomics and Fiber

From an economic point of view, the most agronomically important traits of cotton are the quantity and quality of fiber produced. Cotton fiber is a single-celled, epidermal, ovular trichome with a well-established developmental program (Haigler et al. 2009, 2012) that is largely conserved, but whose variations produce the remarkable differences in fiber between cotton species and between wild and domesticated accessions (Applequist et al. 2001). *Gossypium barbadense* fiber, for example, has an extended elongation phase resulting in the longer, finer phenotype for which

Egyptian cotton is famous (Tu et al. 2007). Conversely, subtle changes in the developmental program for the diploid cultivars result in fiber that is shorter, coarser, and weaker, with less convolution or "spinnability," thereby reducing commercial value (Haigler et al. 2005).

Generally, cotton fiber develops over a 40-day period, during which the fiber first elongates (primary cell wall synthesis) and subsequently undergoes cell wall thickening (secondary cell wall synthesis) until maturity. The pace and duration of each phase of cotton fiber development are crucial for various aspects of fiber quality, such as length (elongation phase) and spinnability (cell wall thickening). Accordingly, research has been focused on characterize either the specific genes or specific developmental time points relevant to fiber development.

Decades of cell biology have characterized many aspects of fiber developmental biology. In particular, much is now known about fiber cell wall patterning, deposition, and composition in wild versus domesticated species, as well as many other aspects (reviewed by Fang et al. 2018; Fang 2018). A series of master transcriptional factors involved in regulation of cotton fiber development, such as MYBs (Shangguan et al. 2008; Walford et al. 2011; Wan et al. 2016; Tan et al. 2016; Wu et al. 2018), HOX3 (Shan et al. 2014; Rombolá-Caldentey et al. 2014), and PRE1 (Zhao et al. 2018), have been reported. Genes such as cellulose synthase (Tuttle et al. 2015; Nixon et al. 2016; Cho et al. 2017) and profilin (Wang et al. 2010; Bao et al. 2011; Argiriou et al. 2012; Gallagher et al. 2016) have been the subject of comparative analyses to determine their roles in fiber development. In other cases, while the genes and/or mechanisms of control are not yet identified, the accumulated results from molecular biological research suggest important roles for hormones (Chen et al. 1997; Dasani and Thaker 2006; Liao et al. 2010; Zhang et al. 2017b), such as ethylene (Shi et al. 2006) and auxin (Singh et al. 2009), and for mediators of sugar flux (Delmer and Haigler 2002; Ruan et al. 2003; Shi et al. 2006) within the developing fiber.

While fiber development has been studied extensively, most research on the *controls* of fiber development is focused on *G. hirsutum*. QTL analysis of a *G. hirsutum* F_2 cross showed that many fiber QTLs were located on the D-genome. While this is perhaps a surprising result, as D-genome *Gossypium* does not produce spinnable fiber (Grover et al. 2019b), it also suggests that factors contributing to the special properties of polyploid cotton may be found in D-genome species, pointing to a potential avenue of improvement via introgression of genes from these species (Jiang et al. 1998; Paterson et al. 2003; Lacape et al. 2005; Han et al. 2006; Rong et al. 2007; Qin et al. 2008; Said et al. 2015). Regulation of fiber-related genes has also been studied in the context of fiber coexpression networks, whose modules tend to exhibit D-genome bias although individual modules may experience the converse (Gallagher et al. 2020). Notably, the A- and D-homoeologous networks are generally similar in structure, despite the lack of spinnable fiber in the D-genome cotton species, suggesting that the foundation exists in those species. The most significant change, however, in the cotton fiber network is the increased co-regulation among genes and homoeologs that results in more tightly regulated gene networks in

domesticated cotton (Gallagher et al. 2020). Network analysis could provide insights into nongenetic avenues of improvement for cotton, or means of altering entire networks within the fiber by editing transcription factors or other "hub" genes, or perhaps by altering their expression levels (Gallagher et al. 2016). It is an exciting prospect that the genomic and technological resources needed to affect these new insights are now becoming mature.

Historically, genetic modification and even breeding of cotton have been challenging due to its polyploid nature and general recalcitrance to regeneration after transformation. Despite this, strides have been made in these areas, with new technologies constantly developing and improving. Diploid cotton is a rich source of variation that could be used to improve any number of traits in commercial cotton, and in some cases this has occurred in the wild (Wendel and Cronn 2003). Because *G. hirsutum* and *G. barbadense* are allopolyploid, introgressing genetic material from diploid species is challenging, although there has been some success historically through hexaploid bridging (Becerra Lopez-Lavalle and Brubaker 2007). Producing knockout or knockdown lines also requires thoughtful consideration to insure both homoeologs are targeted, but here there has also been recent success using virus-induced gene silencing (VIGS) to silence genes in polyploid cotton (Idris et al. 2010; Pang et al. 2013). Perhaps the most challenging aspect of genetic modification in cotton are the hurdles of regeneration after transformation (Juturu et al. 2015). Traditionally, success has been limited to particular genotypes of *G. hirsutum* (Jin et al. 2006, 2012); however, many of the challenges and recent developments in cotton transformation have recently been outlined (Juturu et al. 2015). Some work has been done on ovule culture transformation and with floral dip transformation leading to single-gene transformants (Johar Campus 2005; Zhang and Chen 2012), and strides in large-scale genetic modification are being made, particularly with CRISPR technology (Chen et al. 2017a; Li et al. 2017; Janga et al. 2017; Wang et al. 2018b). New transformation technologies have also shown promise, as in the case of pollen magnetofection (Zhao et al. 2017). These recent advances involving genetic engineering in cotton provide exciting prospects for using the substantial molecular research in *Gossypium* in breeding programs and in generating genetically modified cultivars.

2.11 Polyploidy, Stress Tolerance, and Implications for Crop Improvement in a Changing Environment

As is the case for *G. hirsutum* and *G. barbadense*, many important crops are contemporary polyploids, including wheat, potato, oats, oilseed rape, strawberries, banana, kiwifruit, watermelon, and many others. In these cases, physiological and morphological traits are often distinct in polyploid plants relative to their diploid progenitors. At the cellular level, the increment in cell size with the increase of ploidy level is referred to as "gigas" effect (Müntzing 1936; Stebbins et al. 1971).

Correlated with larger cell size, polyploid individuals often are transgressive relative to their diploid conspecifics at the tissue or organ level, with larger structures of roots, leaves, flowers, and seeds (Stebbins 1950; Dudits et al. 2016). Many of these scaling and stoichiometric issues are poorly understood, as detailed in the lengthy and insightful review by Doyle and Coate (2019). In natural settings, polyploidy often is associated with ecologically marginal areas or niche expansion due to their adaptive capacity to adverse or extreme environments (Lewis 1979; Ehrendorfer 1980; Stebbins 1985; Otto and Whitton 2000; Ni et al. 2009; McIntyre 2012). One generalization that has emerged from these and other papers is that in some environments and at certain periods, polyploids have a fitness advantage relative to their diploid cousins (Leitch and Leitch 2008; Ainouche and Wendel 2014; Soltis et al. 2016; de Peer et al. 2017). This is often suggested to be due to the increased "buffering" capacity afforded by duplicated genes and enhanced vigor from the "fixed" heterozygosity (two homoeologs providing a minimum of two different alleles), together with a diverse suite of genetic and genomic interactions, as discussed above. This remarkable property of polyploids has long attracted the attention of plant breeders, who often have used artificial polyploids as a tool for crop improvement (reviewed in Sattler et al. 2016). As mentioned above, the extraordinary advances in genomic resources combined with gene editing technologies such as CRISPR/Cas9 promise an entirely new opportunity to harness the phenotypic and ecological potentiality of polyploid genomes for specific breeding objectives.

In the case of *Gossypium*, perhaps the most notable ecological consequence of allopolyploidy is that it led to an apparent invasion of new ecological niches. In contrast to diploid species that mostly are found inland, away from coastal margins, allopolyploid species in their native habitats typically occur in coastal and insular environments, which, as pointed out by Fryxell (1965, 1979), is associated with the oceanic dispersal of salt-water-tolerant seeds in the newly evolved allopolyploid lineage. This raises the prospect that allopolyploid cotton inherently has a higher tolerance to salt stress. Recently, a diverse panel of eight diploid (from A- and D-genome groups) and four allopolyploid cotton species were screened for their morphophysiological responses to different levels of salinity stress in hydroponics (Dong et al. 2020). Among these species, the Brazilian allopolyploid *G. mustelinum* exhibits the highest level of salt tolerance; however, no significant difference was noted in allopolyploids relative to either A- or D-genome diploids, due to a wide range of variations within each genome group. Although these observations do not support the speculation of higher salt tolerance in polyploid cotton, there is insufficient evidence to reject this hypothesis because salt stress affects all cotton developmental stages and only the seeding stage was surveyed in this study. Moreover, hydroponic evaluation for salinity tolerance cannot account for the temporal and spatial variations in soil physical and chemical properties under field conditions or in natural settings, which are linked to other types of environmental constraints such as drought and mineral nutrient deficiency. By this means, perhaps, the ecological innovations of allopolyploid cotton are beyond salt tolerance alone, perhaps entailing an increased level of phenotypic plasticity to cope with altered environmental conditions. Other polyploids have provided some empirical support for this notion,

for example, in spotted knapweed (*Centaurea stoebe* L); as compared to the diploid *C. stoebe*, the tetraploids exhibit a generally broader ecological tolerance which may have contributed to their success as invasive species in novel environments (Hahn et al. 2012).

In the context of climate change, ever-increasing variability in weather patterns is threatening cotton production globally due to temperature extremes, irregular rainfalls, and drought, which collectively comprise novel abiotic and biotic stresses. Therefore, it is a pressing matter to develop climate-resilient and stress-tolerant crops to sustain world cotton production. Although some progress has been made in addressing the physiological and molecular aspects of abiotic stress tolerance in cotton (Khan et al. 2018; Zafar et al. 2018; Harkess 2018; Abdelraheem et al. 2019; Chen et al. 2019, 2020), to date this understanding has not led to comparable advances in breeding. To the best of our knowledge, no cotton cultivars with a markedly enhanced abiotic stress tolerance, with high yield and fiber quality, are commercially available (Abdelraheem et al. 2019). This likely reflects the complex genetic basis for these traits, limited resources from the core germplasm pool, and difficulties in incorporated complex physiological capacities into breeding programs. While powerful new transgenic and gene editing methods are beginning to revolutionize technical aspects developing new cotton varieties, it seems likely that the knowledge foundation for this important endeavor will continue to emerge from our understanding of the extraordinary natural diversity in *Gossypium* and the still mysterious consequences of allopolyploid formation.

References

Abdalla AM, Reddy OUK, El-Zik KM, Pepper AE (2001) Genetic diversity and relationships of diploid and tetraploid cottons revealed using AFLP. *Theor Appl Genet* 102:222–229

Abdellatif KF, Khidr YA, El-Mansy YM et al (2012) Molecular diversity of Egyptian cotton (*Gossypium barbadense* L.) and its relation to varietal development. *J Crop Sci Biotechnol* 15:93–99

Abdelraheem A, Esmaeili N et al (2019) Progress and perspective on drought and salt stress tolerance in cotton. *Ind Crop Prod* 130:118–129

Abdul Kadir ZB (1976) DNA evolution in the genus *Gossypium*. *Chromosoma* 56:85–94

Abraham P et al (1940) Cytological studies in Gossypium: chromosome behaviour in the interspecific hybrid *G. arboreum* x *G. stocksii*. *Indian J Agric Sci* 10:285–298

Adams KL, Wendel JF (2013) Dynamics of duplicated gene expression in polyploid cotton. In: Polyploid and hybrid genomics. pp 187–194

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

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

Adams KL, Flagel L, Wendel JF (2009) Responses of the cotton genome to polyploidy. In: Paterson AH (ed) Plant Genetics and Genomics: Crops and Models. Springer, New York, NY, pp 419–4

Ågren JA, Huang H-R, Wright SI (2016) Transposable element evolution in the allotetraploid *Capsella bursa-pastoris*. *Am J Bot* 103:1197–1202

Ahmad MQ, Khan SH, Azhar FM (2012) Decreasing level of genetic diversity in germplasm and cultivars of upland cotton (*Gossypium hirsutum*) in Pakistan. *J Agric Soc Sci* 8:92–96

Ainouche ML, Wendel JF (2014) Polyploid speciation and genome evolution: lessons from recent allopolyploids. In: Pontarotti P (ed) Evolutionary biology: genome evolution, speciation, coevolution and origin of life. Springer International Publishing, Cham, pp 87–113

Alvarez I, Wendel JF (2006) Cryptic interspecific introgression and genetic differentiation within *Gossypium aridum* (Malvaceae) and its relatives. *Evolution* 60:505–517

An Z, Tang Z, Ma B et al (2014) Transposon variation by order during allopolyploidisation between *Brassica oleracea* and *Brassica rapa*. *Plant Biol* 16:825–835

Applequist WL, Cronn R, Wendel JF (2001) Comparative development of fiber in wild and cultivated cotton. *Evol Dev* 3:3–17

Argiriou A, Kalivas A, Michailidis G, Tsaftaris A (2012) Characterization of *PROFILIN* genes from allotetraploid (*Gossypium hirsutum*) cotton and its diploid progenitors and expression analysis in cotton genotypes differing in fiber characteristics. *Mol Biol Rep* 39:3523–3532

Bao Y, Hu G, Flagel LE et al (2011) Parallel up-regulation of the profilin gene family following independent domestication of diploid and allopolyploid cotton (*Gossypium*). *Proc Natl Acad Sci U S A* 108:21152–21157

Basu AK (1996) Current genetic research in cotton in India. *Genetica* 97:279–290

Beasley JO (1940) The origin of American tetraploid *Gossypium* species. *Am Nat* 74:285–286

Beasley JO (1942) Meiotic chromosome behavior in species, species hybrids, haploids, and induced polyploids of *Gossypium*. *Genetics* 27:25–54

Becerra Lopez-Lavalle LA, Brubaker CL (2007) Frequency and fidelity of alien chromosome transmission in *Gossypium* hexaploid bridging populations. *Genome* 50:479–491

Bell AA (1984) Morphology, chemistry, and genetics of *Gossypium* adaptations to pests. In: Timmermann BN, Steelink C, Loewus FA (eds) Phytochemical adaptations to stress. Springer US, Boston, pp 197–230

Bell A, Robinson AF (2004) Development and characteristics of triple species hybrids used to transfer reniform nematode resistance from *Gossypium longicalyx* to *Gossypium hirsutum*. In: Proceedings of the beltwide cotton conferences. pp 422–426

Bell AA, Forest Robinson A, Quintana J et al (2014) Registration of LONREN-1 and LONREN-2 germplasm lines of upland cotton resistant to Reniform Nematode. *J Plant Regist* 8:187

Bell AA, Robinson AF, Quintana J et al (2015) Registration of BARBREN-713 germplasm line of upland cotton resistant to Reniform and Root-Knot Nematodes. *J Plant Regist* 9:89

Ben-David S, Yaakov B, Kashkush K (2013) Genome-wide analysis of short interspersed nuclear elements SINES revealed high sequence conservation, gene association and retrotranspositional activity in wheat. *Plant J* 76:201–210

Boopathi NM, Gopikrishnan A, Selvam NJ et al (2008) Genetic diversity assessment of *G. barbadense* accessions to widen cotton (*Gossypium* spp.) gene pool for improved fibre quality. *J Cotton Res Dev* 22:135–138

Bowman DT, McCarty JC (1997) Thrips (Thysanoptera: Thripidae) Tolerance in cotton: sources and heritability. *J Entomol Sci* 32:460–471

Brown HB, Ware JO (1958) Cotton. McGraw-Hill, New York

Brubaker CL, Wendel JF (1994) Reevaluating the origin of domesticated cotton (*Gossypium hirsutum*; Malvaceae) using nuclear restriction fragment length polymorphisms (RFLPs). *Am J Bot* 81:1309–1326

Brubaker CL, Wendel JF (2001) RFLP diversity in cotton. In: Jenkins JN, Saha S (eds) Genetic improvement of cotton: emerging technologies. Science Publishers, Enfield, pp 81–102

Brubaker CL, Koontz JA, Wendel JF (1993) Bidirectional cytoplasmic and nuclear introgression in the New World cottons *Gossypium barbadense* and *G. hirsutum* (Malvaceae). *Am J Bot* 80:1203–1208

Brubaker CL, Benson CG, Miller C, Leach DN (1996) Occurrence of terpenoid aldehydes and lysigenous cavities in the “glandless” seeds of Australian *Gossypium* species. *Aust J Bot* 44:601

Brubaker CL, Borland FM, Wendel JF (1999a) The origin and domestication of cotton. In: Smith CW, Cothren JT (eds) Cotton: origin, history, technology and production. Wiley, New York, pp 3–31

Brubaker CL, Paterson AH, Wendel JF (1999b) Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. *Genome* 42:184–203

Bruggmann R, Bharti AK, Gundlach H et al (2006) Uneven chromosome contraction and expansion in the maize genome. *Genome Res* 16:1241–1251

Cao Z, Zhu X, Chen H, Zhang T (2015) Fine mapping of clustered quantitative trait loci for fiber quality on chromosome 7 using a *Gossypium barbadense* introgressed line. *Mol Breed* 35:215

Chaudhary B, Flagel L, Stupar RM et al (2009) Reciprocal Silencing, Transcriptional Bias and Functional Divergence of Homeologs in Polyploid Cotton (*Gossypium*). *Genetics* 182:503–517

Chee PW, Paterson AH, Udall JA, Wendel JF (2016) Interspecific hybridization for upland cotton improvement. In: Mason AS (ed) Polyploidy and hybridization for crop improvement. pp 1–20

Chen ZJ (2007) Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annu Rev Plant Biol* 58:377–406

Chen J-G, Du X-M, Zhou X, Zhao H-Y (1997) Levels of cytokinins in the ovules of cotton mutants with altered fiber development. *J Plant Growth Regul* 16:181–185

Chen Z, Feng K, Grover CE et al (2016) Chloroplast DNA structural variation, phylogeny, and age of divergence among diploid cotton species. *PLoS One* 11:e0157183

Chen X, Lu X, Shu N et al (2017a) Targeted mutagenesis in cotton (*Gossypium hirsutum* L.) using the CRISPR/Cas9 system. Sci Rep 7:44304. https://doi.org/10.1038/srep44304

Chen Z, Grover CE, Li P et al (2017b) Molecular evolution of the plastid genome during diversification of the cotton genus. *Mol Phylogenet Evol* 112:268–276

Chen Z, Nie H, Grover CE et al (2017c) Entire nucleotide sequences of *Gossypium raimondii* and *G. arboreum* mitochondrial genomes revealed A-genome species as cytoplasmic donor of the allotetraploid species. *Plant Biology* 19:484–493

Chen Z, Zhao N, Li S et al (2017d) Plant mitochondrial genome evolution and cytoplasmic male sterility. *CRC Crit Rev Plant Sci* 36:55–69

Chen H, Li Y, Ma X et al (2019) Analysis of potential strategies for cadmium stress tolerance revealed by transcriptome analysis of upland cotton. *Sci Rep* 9:86

Chen ZJ, Sreedasyam A, Ando A, et al (2020) Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement. *Nat Genet* 52:525–533

Cheng F, Wu J, Fang L et al (2012) Biased gene fractionation and dominant gene expression among the subgenomes of *Brassica rapa*. *PLoS One* 7:e36442

Cho SH, Purushotham P, Fang C et al (2017) Synthesis and self-assembly of cellulose microfibrils from reconstituted cellulose synthase. *Plant Physiol* 175:146–156

Conover JL, Karimi N, Stenz N et al (2019) A Malvaceae mystery: a mallow maelstrom of genome multiplications and maybe misleading methods? *J Integr Plant Biol* 61:12–31

Coppens d'Eeckenbrugge G, 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

Craven LA, Stewart JMD, Brown AHD, Grace JP (1994) The Australian wild species of *Gossypium*. In: Challenging the future: proceedings of the world cotton research conference I. CSIRO. pp 278–281

Cronn R, Wendel JF (2003) Cryptic trysts, genomic mergers, and plant speciation: research review. *New Phytol* 161:133–142

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

Cronn RC, Small RL, Wendel JF (1999) Duplicated genes evolve independently after polyploid formation in cotton. *Proc Natl Acad Sci* 96:14406–14411

Cronn RC, Small RL, Haselkorn T, Wendel JF (2002) Rapid diversification of the cotton genus (*Gossypium*: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. *Am J Bot* 89:707–725

Cronn R, Small RL, Haselkorn T, Wendel JF (2003) Cryptic repeated genomic recombination during speciation in *Gossypium gossypioides*. *Evolution* 57:2475–2489
Dasani SH, Thaker VS (2006) Role of abscisic acid in cotton fiber development. *Russ J Plant Physiol* 53:62–67
Davie JH, Hugh Davie J (1933) Cytological studies in the Malvaceae and certain related families. *J Genet* 28:33–67
de Peer YV, Van de Peer Y, Mizrachi E, Marchal K (2017) The evolutionary significance of polyploidy. *Nat Rev Genet* 18:411–424
DeJoode DR, Wendel JF (1992) Genetic diversity and origin of the Hawaiian Islands cotton, *Gossypium tomentosum*. *Am J Bot* 79:1311–1319
Delmer DP, Haigler CH (2002) The regulation of metabolic flux to cellulose, a major sink for carbon in plants. *Metab Eng* 4:22–28
Dempewolf H, Baute G, Anderson J et al (2017) Past and future use of wild relatives in crop breeding. *Crop Sci* 57:1070–1082
Ding M, Chen ZJ (2018) Epigenetic perspectives on the evolution and domestication of polyploid plant and crops. *Curr Opin Plant Biol* 42:37–48
Dong Y, Hu G, Yu J, et al (2020) Salt-tolerance diversity in diploid and polyploid cotton (*Gossypium*) species. *Plant J* 101:1135–1151
Doyle JJ, Coate JE (2019) Polyploidy, the nucleotype, and novelty: the impact of genome doubling on the biology of the cell. *Int J Plant Sci* 180:1–52
Du X, Huang G, He S et al (2018) Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits. *Nat Genet* 50:796–802
Dudits D, Török K, Cseri A et al (2016) Response of organ structure and physiology to autotetraploidization in early development of energy willow *Salix viminalis* L. *Plant Physiol* 170(3):1504–1523. https://doi.org/10.1104/pp.15.01679
Ehrendorfer F (1980) Polyploidy and distribution. In: Polyploidy. pp 45–60
Endrizzi JE, Turcotte EL, Kohel RJ (1985) Genetics, cytology, and evolution of *Gossypium*. In: Caspari EW, Scandalios JG (eds) Advances in genetics. Academic, New York, pp 271–375
Endrizzi JE, Katterman FRH, Geever RF (Department of Plant Sciences, University of Arizona, Tucson, AZ 85721 (USA)) (1989) DNA hybridization and time of origin of three species of *Gossypium*. In: Evolutionary trends in plants (United Kingdom). p 3
Fan HH, Wei J, Li TC et al (2013) DNA methylation alterations of upland cotton (*Gossypium hirsutum*) in response to cold stress. *Acta Physiol Plant* 35:2445–2453
Fan X, Guo Q, Xu P et al (2015) Transcriptome-wide identification of salt-responsive members of the *WRKY* gene family in *Gossypium aridum*. *PLoS One* 10:e0126148
Fang DD (2018) Cotton fiber: physics, chemistry and biology. Springer, New York
Fang DD, Hinze LL, Percy RG et al (2013) A microsatellite-based genome-wide analysis of genetic diversity and linkage disequilibrium in Upland cotton (*Gossypium hirsutum* L.) cultivars from major cotton-growing countries. *Euphytica* 191:391–401
Fang L, Gong H, Hu Y et al (2017a) Genomic insights into divergence and dual domestication of cultivated allotetraploid cottons. *Genome Biol* 18:33
Fang L, Wang Q, Hu Y et al (2017b) Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. *Nat Genet* 49:1089–1098
Fang DD, Naoumkina M, Kim HJ (2018) Unraveling cotton fiber development using fiber mutants in the post-genomic era. *Crop Sci* 58:2214–2228
Feldman M, Liu B, Segal G et al (1997) Rapid elimination of low-copy DNA sequences in polyploid wheat: a possible mechanism for differentiation of homoeologous chromosomes. *Genetics* 147:1381–1387
Feng C, Ulloa M, Perez-M C, Stewart JM (2011) Distribution and molecular diversity of arborescent *Gossypium* species. *Botany* 89:615–624
Flagel LE, Wendel JF (2010) Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation. *New Phytol* 186:184–193

Flagel L, Udall J, Nettleton D, Wendel J (2008) Duplicate gene expression in allopolyploid *Gossypium* reveals two temporally distinct phases of expression evolution. *BMC Biol* 6:16

Flagel LE, Wendel JF, Udall JA (2012) Duplicate gene evolution, homoeologous recombination, and transcriptome characterization in allopolyploid cotton. *BMC Genomics* 13:302

Freeling M, Thomas BC (2006) Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. *Genome Res* 16:805–814

Fryxell PA (1965) A revision of the Australian species of *Gossypium* with observations on the occurrence of *Thespesia* in Australia (Malvaceae). *Aust J Bot* 13:71–102

Fryxell PA (1968) A redefinition of the tribe Gossypieae. *Bot Gaz* 129:296–308

Fryxell PA (1979) The natural history of the cotton tribe (Malvaceae, tribe Gossypieae). TAMU Press, College Station

Fryxell PA (1986) Ecological adaptations of *Gossypium* species. In: Cotton Physiology. The Cotton Foundation, Memphis, TN, pp 1–7

Fryxell PA (1992) A revised taxonomic interpretation of *Gossypium* L (Malvaceae). Rheeda 2:108–165

Fryxell PA, Kohel RJ, Lewis CF (1984) Taxonomy and germplasm resources. In: Cotton. American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, pp 27–57

Fryxell PA, Koch SD, Koch SD (1987) New or noteworthy species of flowering plants from the Sierra Madre de Sur of Guerrero and Michoacán, Mexico. *J Syst Evol Bot* 11:539–561

Galau GA, Wilkins TA (1989) Alloplasmic male sterility in AD allotetraploid *Gossypium hirsutum* upon replacement of its resident A cytoplasm with that of D species *G. harknessii*. *Theor Appl Genet* 78:23–30

Gallagher JP, Grover CE, Hu G, Wendel JF (2016) Insights into the ecology and evolution of polyploid plants through network analysis. *Mol Ecol* 25:2644–2660

Gallagher JP, Grover CE, Rex K et al (2017) A new species of cotton from wake atoll, *Gossypium stephensii* (Malvaceae). *Syst Bot* 42:115–123

Gallagher JP, Grover CE, Hu G, et al (2020) Conservation and divergence in duplicated fiber coexpression networks accompanying domestication of the polyploid *Gossypium hirsutum* L. *G3* 10:2879–2892

Gao W, Long L, Tian X et al (2017) Genome editing in cotton with the CRISPR/Cas9 system. *Front Plant Sci* 8:1364

Gerstel DU (1953) Chromosomal translocations in interspecific hybrids of the genus *Gossypium*. *Evolution* 7:234–244

Grover CE, Yu Y, Wing RA et al (2008) A phylogenetic analysis of indel dynamics in the cotton genus. *Mol Biol Evol* 25:1415–1428

Grover CE, Gallagher JP, Szadkowski EP et al (2012a) Homoeolog expression bias and expression level dominance in allopolyploids. *New Phytol* 196:966–971

Grover CE, Salmon A, Wendel JF (2012b) Targeted sequence capture as a powerful tool for evolutionary analysis1. *Am J Bot* 99:312–319

Grover CE, Gallagher JP, Jareczek JJ et al (2015a) Re-evaluating the phylogeny of allopolyploid *Gossypium* L. *Mol Phylogenet Evol* 92:45–52

Grover CE, Gallagher JP, Wendel JF (2015b) Candidate gene identification of flowering time genes in cotton. *Plant Genome* 8: plantgenome2014.12.0098

Grover CE, Zhu X, Grupp KK et al (2015c) Molecular confirmation of species status for the allopolyploid cotton species, *Gossypium ekmanianum* Wittmack. *Genet Resour Crop Evol* 62:103–114

Grover CE, Arick MA 2nd, Conover JL et al (2017a) Comparative genomics of an unusual biogeographic disjunction in the cotton tribe (*Gossypieae*) yields insights into genome downsizing. *Genome Biol Evol* 9:3328–3344

Grover CE, Gallagher JP, Szadkowski EP et al (2017b) Nucleotide diversity in the two co-resident genomes of allopolyploid cotton. *Plant Syst Evol* 303:1021–1042

Grover CE, Arick MA 2nd, Thrash A et al (2019a) Insights into the evolution of the New World diploid cottons (*Gossypium*, Subgenus Houzingenia) based on genome sequencing. *Genome Biol Evol* 11:53–71

Grover CE, Yoo M-J, Gore MA et al (2019b) Genetic analysis of the transition from wild to domesticated cotton (*G. hirsutum*). bioRxiv 616763

Guan X, Pang M, Nah G et al (2014) miR828 and miR858 regulate homoeologous *MYB2* gene functions in *Arabidopsis* trichome and cotton fibre development. *Nat Commun* 5:3050

Gulati AN, Turner AJ (1928) A note on the early history of cotton. Indian Central Cotton Committee, Technological Laboratory

Guo W-Z, Zhou B-L, Yang L-M et al (2006) Genetic diversity of landraces in *Gossypium arboreum* L. race sinense assessed with simple sequence repeat markers. *J Integr Plant Biol* 48:1008–1017

Guo H, Wang X, Gundlach H et al (2014) Extensive and biased intergenomic nonreciprocal DNA exchanges shaped a nascent polyploid genome, *Gossypium* (Cotton). *Genetics* 197:1153–1163

Guseinov VA, Kiryanov GI, Vanyushin BF (1975) Intragenome distribution of 5-methylcytosine in DNA of healthy and wilt-infected cotton plants (*Gossypium hirsutum* L.). *Mol Biol Rep* 2:59–63

Hahn MA, van Kleunen M, Müller-Schärer H (2012) Increased phenotypic plasticity to climate may have boosted the invasion success of polyploid Centaurea stoebe. *PLoS One* 7:e50284

Haigler CH, Zhang D, Wilkerson CG (2005) Biotechnological improvement of cotton fibre maturity. *Physiol Plant* 124:285–294

Haigler CH, Singh B, Wang G, Zhang D (2009) Genomics of cotton fiber secondary wall deposition and cellulose biogenesis. In: Paterson AH (ed) Genetics and genomics of cotton. Springer US, New York, pp 385–417

Haigler CH, Betancur L, Stiff MR, Tuttle JR (2012) Cotton fiber: a powerful single-cell model for cell wall and cellulose research. *Front Plant Sci* 3:104

Han Z, Wang C, Song X et al (2006) Characteristics, development and mapping of *Gossypium hirsutum* derived EST-SSRs in allotetraploid cotton. *Theor Appl Genet* 112:430–439

Hanson R, Zhao XP, Islam-Faridi MN et al (1998) Evolution of interspersed repetitive elements in *Gossypium* (Malvaceae). *Am J Bot* 85:1364–1368

Hanson RE, Islam-Faridi MN, Crane CF et al (2000) *Ty1-copia*-retrotransposon behavior in a polyploid cotton. *Chromosome Res* 8:73–76

Harkess A (2018) Handling the heat: methylome variation underlying heat tolerance in cotton. *Plant Cell* 30:1947–1948

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

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

Hendrix B, Stewart JM (2005) Estimation of the nuclear DNA content of *Gossypium* species. *Ann Bot 95*:789–797

Hinze LL, Fang DD, Gore MA et al (2015) Molecular characterization of the *Gossypium* diversity reference set of the US National Cotton Germplasm Collection. *Theor Appl Genet* 128:313–327

Hinze LL, Gazave E, Gore MA et al (2016) Genetic diversity of the two commercial tetraploid cotton species in the *Gossypium* diversity deference set. *J Hered* 107:274–286

Hovav R, Udall JA, Chaudhary B et al (2008) Partitioned expression of duplicated genes during development and evolution of a single cell in a polyploid plant. *Proc Natl Acad Sci U S A* 105:6191–6195

Hovav R, Faigenboim-Doron A, Kadmon N et al (2015) A transcriptome profile for developing seed of polyploid cotton. *Plant Genome* 8

Hu G, Wendel JF (2019) Cis-trans controls and regulatory novelty accompanying allopolyploidization. *New Phytol* 221:1691–1700

Hu G, Hawkins JS, Grover CE, Wendel JF (2010) The history and disposition of transposable elements in polyploid *Gossypium*. *Genome* 53:599–607

Hu G, Houston NL, Pathak D et al (2011) Genomically biased accumulation of seed storage proteins in allopolyploid cotton. *Genetics* 189:1103–1115

Hu G, Koh J, Yoo M-J et al (2013) Proteomic profiling of developing cotton fibers from wild and domesticated *Gossypium barbadense*. *New Phytol* 200:570–582
Hu G, Koh J, Yoo M-J et al (2015) Gene-expression novelty in allopolyploid cotton: a proteomic perspective. *Genetics* 200:91–104
Hu G, Hovav R, Grover CE et al (2016) Evolutionary conservation and divergence of gene coexpression networks in *Gossypium* (cotton) seeds. *Genome Biol Evol* 8:3765–3783
Hu Y, Chen J, Fang L et al (2019) *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. *Nat Genet* 51(4):739–748. https://doi.org/10.1038/s41588-019-0371-5
Hutchinson JB (1951) Intra-specific differentiation in *Gossypium hirsutum*. *Heredity* 5:161–193
Hutchinson JB (1954) New evidence on the origin of the old world cottons. *Heredity* 8:225–241
Hutchinson JB (1959) The application of genetics to cotton improvement, 1st edn. Cambridge University Press, Cambridge
Hutchinson JB, Silow RA, Stephens SG (1947) The evolution of *Gossypium* and the differentiation of the cultivated cottons. Geoffrey Cumberlege, Oxford University Press, London
Idris AM, Tuttle JR, Robertson D et al (2010) Differential Cotton leaf crumple virus-VIGS-mediated gene silencing and viral genome localization in different *Gossypium hirsutum* genetic backgrounds. *Physiol Mol Plant Pathol* 75:13–22
Iqbal MJ, Aziz N, Saeed NA et al (1997) Genetic diversity evaluation of some elite cotton varieties by RAPD analysis. *Theor Appl Genet* 94:139–144
Iqbal MJ, Reddy OUK, El-Zik KM, Pepper AE (2001) A genetic bottleneck in the 'evolution under domestication' of upland cotton *Gossypium hirsutum* L. examined using DNA fingerprinting. *Theor Appl Genet* 103:547–554
Janga MR, Campbell LM, Rathore KS (2017) CRISPR/Cas9-mediated targeted mutagenesis in upland cotton (*Gossypium hirsutum* L.). *Plant Mol Biol* 94:349–360
Jena SN, Srivastava A, Singh UM et al (2012) Analysis of genetic diversity, population structure and linkage disequilibrium in elite cotton (*Gossypium* L.) germplasm in India. *Crop Pasture Sci* 62:859–875
Jiang C, Wright RJ, El-Zik KM, Paterson AH (1998) Polyploid formation created unique avenues for response to selection in *Gossypium* (cotton). *Proc Natl Acad Sci U S A* 95:4419–4424
Jiang CX, Chee PW, Draye X et al (2000) Multilocus interactions restrict gene introgression in interspecific populations of polyploid *Gossypium* (cotton). *Evolution* 54:798–814
Jiao Y, Wickett NJ, Ayyampalayam S et al (2011) Ancestral polyploidy in seed plants and angiosperms. *Nature* 473:97–100
Jin S, Zhang X, Nie Y et al (2006) Identification of a novel elite genotype for in vitro culture and genetic transformation of cotton. *Biol Plant* 50:519–524
Jin S-X, Liu G-Z, Zhu H-G et al (2012) Transformation of upland cotton (*Gossypium hirsutum* L.) with *gfp* gene as a visual marker. *J Integr Agric* 11:910–919
Jin X, Pang Y, Jia F et al (2013) A potential role for *CHH* DNA methylation in cotton fiber growth patterns. *PLoS One* 8:e60547
Johar Campus H (2005) In-ovule embryo culture: a novel method of cotton transformation. *Pak J Biol Sci* 8:297–301
Johnson BL (1975) *Gossypium palmeri* and a polyphyletic origin of the New World cottons. *Bull Torrey Bot Club* 102:340–349
Juturu VN, Mekala GK, Kirti PB (2015) Current status of tissue culture and genetic transformation research in cotton (*Gossypium* spp.). *Plant Cell Tissue Organ Cult* 120:813–839
Kashkush K, Feldman M, Levy AA (2003) Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nat Genet* 33:102–106
Katterman FRH, Ergle DR (1970) A study of quantitative variations of nucleic acids in *Gossypium*. *Phytochemistry* 9:2007–2010
Kawakami T, Strakosh SC, Zhen Y, Ungerer MC (2010) Different scales of *Ty1/copia*-like retrotransposon proliferation in the genomes of three diploid hybrid sunflower species. *Heredity* 104:341–350
Keyte AL, Percifield R, Liu B, Wendel JF (2006) Infraspecific DNA methylation polymorphism in cotton (*Gossypium hirsutum* L.). *J Hered* 97:444–450

Khadi BM, Santhy V, Yadav MS (2010) Cotton: an introduction. In: Cotton: biotechnological advances. Springer, Berlin, Heidelberg, pp 1–14

Khan A, Pan X, Najeeb U et al (2018) Coping with drought: stress and adaptive mechanisms, and management through cultural and molecular alternatives in cotton as vital constituents for plant stress resilience and fitness. *Biol Res* 51:47

Konan NO, Baudoin J-P, D'Hont A, Mergeai G (2009) Bridging classical and molecular cytogenetics of *Gossypium*. In: Paterson AH (ed) Genetics and genomics of cotton. Springer US, New York, pp 257–281

Kranthi KR (2018) Cotton production practices: snippets from global data 2017. The ICAC Recorder XXXVI:4–14

Krapovickas A, Seijo G, Seijo JG (2008) *Gossypium ekmanianum* (Malvaceae), algodon silvestre de la Republica Dominicana. *Bonplandia* 55–63

Lacape J-M, Nguyen T-B, Courtois B et al (2005) QTL analysis of cotton fiber quality using multiple *Gossypium hirsutum Gossypium barbadense* backcross generations. *Crop Sci* 45:123–140

Lacape J-M, Dessauw D, Rajab M et al (2007) Microsatellite diversity in tetraploid *Gossypium* germplasm: assembling a highly informative genotyping set of cotton SSRs. *Mol Breed* 19:45–58

LaDuke JC, Doebley J (1995) A chloroplast DNA based phylogeny of the Malvaceae. *Syst Bot* 20:259

Lawrence WJC (1931) The secondary association of chromosomes. *Cytologia* 2:352–384

Leitch AR, Leitch IJ (2008) Genomic plasticity and the diversity of polyploid plants. *Science* 320:481–483

Lewis WH (1979) Polyploidy in species populations. *Basic Life Sci* 13:103–144

Li C, Zhang B (2019) Genome editing in cotton using CRISPR/Cas9 system. *Methods Mol Biol* 1902:95–104

Li F, Fan G, Wang K et al (2014) Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nat Genet* 46:567–572

Li F, Fan G, Lu C et al (2015) Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nat Biotechnol* 33:524–530

Li C, Unver T, Zhang B (2017) A high-efficiency CRISPR/Cas9 system for targeted mutagenesis in Cotton (*Gossypium hirsutum* L.). *Sci Rep* 7:43902

Li B, Rui H, Li Y et al (2019a) Robust CRISPR/Cpf1(Cas12a) mediated genome editing in allotetraploid cotton (*G. hirsutum*). *Plant Biotechnol J* 17(10):1862–1864. https://doi.org/10.1111/pbi.13147

Li J, Manghwar H, Sun L et al (2019b) Whole genome sequencing reveals rare off-target mutations and considerable inherent genetic or/and somaclonal variations in CRISPR/Cas9-edited cotton plants. *Plant Biotechnol J* 17:858–868

Liao W, Zhang J, Xu N, Peng M (2010) The role of phytohormones in cotton fiber development. *Russ J Plant Physiol* 57:462–468

Lim KY, Kovarik A, Matyasek R et al (2007) Sequence of events leading to near-complete genome turnover in allopolyploid *Nicotiana* within five million years. *New Phytol* 175:756–763

Liu Z, Adams KL (2007) Expression partitioning between genes duplicated by polyploidy under abiotic stress and during organ development. *Curr Biol* 17:1669–1674

Liu B, Vega JM, Feldman M (1998a) Rapid genomic changes in newly synthesized amphiploids of *Triticum* and *Aegilops*. II. Changes in low-copy coding DNA sequences. *Genome* 41:535–542

Liu B, Vega JM, Segal G et al (1998b) Rapid genomic changes in newly synthesized amphiploids of *Triticum* and *Aegilops*. I. Changes in low-copy noncoding DNA sequences. *Genome* 41:272–277

Liu B, Brubaker CL, Mergeai G et al (2001a) Polyploid formation in cotton is not accompanied by rapid genomic changes. *Genome* 44:321–330

Liu Q, Brubaker CL, Green AG et al (2001b) Evolution of the FAD2-1 fatty acid desaturase 5' UTR intron and the molecular systematics of *Gossypium* (Malvaceae). *Am J Bot* 88:92–102

Liu L, Guo W, Zhu X, Zhang T (2003) Inheritance and fine mapping of fertility restoration for cytoplasmic male sterility in *Gossypium hirsutum* L. *Theor Appl Genet* 106:461–469

Liu Q, Chen Y, Wang Y et al (2015a) A new synthetic allotetraploid (A1A1G2G2) between *Gossypium herbaceum* and *G. australe*: bridging for simultaneously transferring favorable genes from these two diploid species into upland cotton. *PLoS One* 10:e0123209

Liu X, Zhao B, Zheng H-J et al (2015b) *Gossypium barbadense* genome sequence provides insight into the evolution of extra-long staple fiber and specialized metabolites. *Sci Rep* 5:14139

Long L, Guo D-D, Gao W et al (2018) Optimization of CRISPR/Cas9 genome editing in cotton by improved sgRNA expression. *Plant Methods* 14:85

Lopes FR, Jjingo D, da Silva CRM et al (2013) Transcriptional activity, chromosomal distribution and expression effects of transposable elements in *Coffea* genomes. *PLoS One* 8:e78931

Lubbers EL, Chee PW (2009) The worldwide gene pool of *G. hirsutum* and its improvement. In: Paterson AH (ed) Genetics and Genomics of Cotton. Springer US, New York, NY, pp 23–52

Lukens LN, Pires JC, Leon E et al (2006) Patterns of sequence loss and cytosine methylation within a population of newly resynthesized *Brassica napus* allopolyploids. *Plant Physiol* 140:336–348

Ma Z, He S, Wang X et al (2018) Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. *Nat Genet* 50:803–813

Madlung A (2013) Polyploidy and its effect on evolutionary success: old questions revisited with new tools. *Heredity* 110:99–104

Madlung A, Tyagi AP, Watson B et al (2005) Genomic changes in synthetic *Arabidopsis* polyploids. *Plant J* 41:221–230

Mammadov J, Buyyarapu R, Guttikonda SK et al (2018) Wild relatives of maize, rice, cotton, and soybean: treasure troves for tolerance to biotic and abiotic stresses. *Front Plant Sci* 9:886

May OL (2001) Registration of PD 94045 germplasm line of Upland cotton. *Crop Sci* 41:279–280

McClintock B (1984) The significance of responses of the genome to challenge. *Science* 226:792–801

McFadden H, Beasley D, Brubaker CL (2004) Assessment of *Gossypium sturtianum* and *G. australe* as potential sources of Fusarium wilt resistance to cotton. *Euphytica* 138:61–72

McIntyre PJ (2012) Polyploidy associated with altered and broader ecological niches in the *Claytonia perfoliata* (Portulacaceae) species complex. *Am J Bot* 99:655–662

Mehetre SS (2010) Wild *Gossypium anomalum*: a unique source of fibre fineness and strength. *Curr Sci* 99:58–71

Menzel MY, Brown MS (1954) The significance of multivalent formation in three-species *Gossypium* hybrids. *Genetics* 39:546–557

Meredith WR (1991) Contributions of introductions to cotton improvement. In: Use of plant introductions in cultivar development part 1. Crop Science Society of America, Madison, pp 127–146

Meshram LD (1994) Development of male sterile system from various sources in cotton (*Gossypium* spp.). *PKV Res J* 18:83–86

Meyer VG (1975) Male sterility from *Gossypium harknessii*. *J Hered* 66:23–27

Meyer JR, Meyer VG (1961) Origin and inheritance of nectariless cotton. *Crop Sci* 1:167

Müntzing A (1936) The evolutionary significance of autopolyploidy. *Hereditas* 21:363–378

Muravenko OV, Fedotov AR, Punina EO et al (1998) Comparison of chromosome BrdU-Hoechst-Giemsa banding patterns of the A1 and (AD)2 genomes of cotton. *Genome* 41:616–625

Nazeer W, Ahmad S, Mahmood K et al (2014) Introgression of genes for cotton leaf curl virus resistance and increased fiber strength from *Gossypium stocksii* into upland cotton (*G. hirsutum*). *Genet Mol Res* 13:1133–1143

Ni Z, Kim E-D, Ha M et al (2009) Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. *Nature* 457:327–331

Nixon BT, Mansouri K, Singh A et al (2016) Comparative structural and computational analysis supports eighteen cellulose synthases in the plant cellulose synthesis complex. *Sci Rep* 6:28696

Oluoch G, Zheng J, Wang X et al (2016) QTL mapping for salt tolerance at seedling stage in the interspecific cross of *Gossypium tomentosum* with *Gossypium hirsutum*. *Euphytica* 209:223–235

Osabe K, Clement JD, Bedon F et al (2014) Genetic and DNA methylation changes in cotton (*Gossypium*) genotypes and tissues. *PLoS One* 9:e86049

Otto SP, Whitton J (2000) Polyploid incidence and evolution. *Annu Rev Genet* 34:401–437

Ozkan H, Levy AA, Feldman M (2001) Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. *Plant Cell* 13:1735–1747

Page JT, Huynh MD, Liechty ZS et al (2013) Insights into the evolution of cotton diploids and polyploids from whole-genome re-sequencing. *G3* 3:1809–1818

Page JT, Liechty ZS, Alexander RH et al (2016) DNA sequence evolution and rare homoeologous conversion in tetraploid cotton. *PLoS Genet* 12:e1006012

Pang J, Zhu Y, Li Q et al (2013) Development of *Agrobacterium*-mediated virus-induced gene silencing and performance evaluation of four marker genes in *Gossypium barbadense*. *PLoS One* 8:e73211

Parisod C, Senerchia N (2012) Responses of transposable elements to polyploidy. In: Grandbastien M-A, Casacuberta JM (eds) Plant transposable elements: impact on genome structure and function. Springer, Berlin, Heidelberg, pp 147–168

Paterson AH, Bowers JE, Burow MD et al (2000) Comparative genomics of plant chromosomes. *Plant Cell* 12:1523

Paterson AH, Saranga Y, Menz M et al (2003) QTL analysis of genotype × environment interactions affecting cotton fiber quality. *Theor Appl Genet* 106:384–396

Paterson AH, Wendel JF, Gundlach H et al (2012) Repeated polyploidization of *Gossypium genomes* and the evolution of spinnable cotton fibres. *Nature* 492:423–427

Paun O, Fay MF, Soltis DE, Chase MW (2007) Genetic and epigenetic alterations after hybridization and genome doubling. *Taxon* 56:649–656

Percy RG, Wendel JF (1990) Allozyme evidence for the origin and diversification of *Gossypium barbadense* L. *Theor Appl Genet* 79:529–542

Phillips LL (1964) Segregation in new allopolyploids of *Gossypium*. V. multivalent formation in New World x Asiatic and New World x Wild American hexaploids. *Am J Bot* 51:324–329

Phillips LL (1966) The cytology and phylogenetics of the diploid species of *Gossypium*. *Am J Bot* 53:328–335

Phillips LL (1977) Interspecific incompatibility in *Gossypium*. Temperature-conditional lethality in hybrids of G. klotzschianum. *Am J Bot* 64:914–915

Phillips LL, Clement D (1967) Variation in the diploid *Gossypium* species of Baja, California. *Madrono* 19:137–147

Phillips LL, Strickland MA (1966) The cytology of a hybrid between *G. hirsutum* and *G. longicalyx*. *Can J Genet Cytol* 8:91–95

Phuphathanaphong L (2006) *Thepparatia* (Malvaceae), a new genus from. *Thailand* 1:195–200

Piednoël M, Carrete-Vega G, Renner SS (2013) Characterization of the LTR retrotransposon repertoire of a plant clade of six diploid and one tetraploid species. *Plant J* 75:699–709

Pinto de Menezes IP, Barroso PAV, Hoffmann LV et al (2010) Genetic diversity of mocó cotton (*Gossypium hirsutum* race marie-galante) from the northeast of Brazil: implications for conservation. *Botany* 88:765–773

Piperno DR, Pearsall DM (1998) Background of tropical agricultural origins. Academic, San Diego

Qin H, Guo W, Zhang Y-M, Zhang T (2008) QTL mapping of yield and fiber traits based on a four-way cross population in *Gossypium hirsutum* L. *Theor Appl Genet* 117:883–894

Rajendran TP, Venugopalan MV, Praharaj CS (2005) Cotton research towards sufficiency to Indian textile industry. *Indian J Agric Sci* 75:699–708

Ram SG, Thiruvengadam V, Ramakrishnan SH, Kannan Bapu JR (2008) Investigation on prezygotic barriers in the interspecific crosses involving *Gossypium barbadense* and four diploid wild species. *Euphytica* 159:241–248

Rambani A, Page JT, Udall JA (2014) Polyploidy and the petal transcriptome of *Gossypium*. *BMC Plant Biol* 14:3

Rapp RA, Udall JA, Wendel JF (2009) Genomic expression dominance in allopolyploids. *BMC Biol* 7:18

Reddy UK, Nimmakayala P, Abburi VL et al (2017) Genome-wide divergence, haplotype distribution and population demographic histories for *Gossypium hirsutum* and *Gossypium barbadense* as revealed by genome-anchored SNPs. *Sci Rep* 7:41285

Reinisch AJ, Dong JM, Brubaker CL et al (1994) A detailed RFLP map of cotton, *Gossypium hirsutum* x *Gossypium barbadense*: chromosome organization and evolution in a disomic polyploid genome. *Genetics* 138:829–847

Renny-Byfield S, Gong L, Gallagher JP, Wendel JF (2015) Persistence of subgenomes in paleopolyploid cotton after 60 my of evolution. *Mol Biol Evol* 32:1063–1071

Renny-Byfield S, Page JT, Udall JA et al (2016) Independent domestication of two Old World cotton species. *Genome Biol Evol* 8:1940–1947

Robinson AF, Bell AA, Dighe ND et al (2007) Introgression of resistance to nematode *Rotylenchulus reniformis* into upland cotton (*Gossypium hirsutum*) from *Gossypium longicalyx*. *Crop Sci* 47:1865–1877

Rombolá-Caldentey B, Rueda-Romero P, Iglesias-Fernández R et al (2014) *Arabidopsis* DELLA and two HD-ZIP transcription factors regulate GA signaling in the epidermis through the L1 box *cis*-element. *Plant Cell* 26:2905–2919

Rong J, Feltus FA, Waghmare VN et al (2007) Meta-analysis of polyploid cotton QTL shows unequal contributions of subgenomes to a complex network of genes and gene clusters implicated in lint fiber development. *Genetics* 176:2577–2588

Rossen J, Dillehay TD, Ugent D (1996) Ancient cultigens or modern intrusions?: evaluating plant remains in an Andean case study. *J Archaeol Sci* 23:391–407

Ruan Y-L, Llewellyn DJ, Furbank RT (2003) Suppression of sucrose synthase gene expression represses cotton fiber cell initiation, elongation, and seed development. *Plant Cell* 15:952–964

Saha S, Jenkins JN, Wu J et al (2006) Effects of chromosome-specific introgression in upland cotton on fiber and agronomic traits. *Genetics* 172:1927–1938

Said JI, Song M, Wang H et al (2015) A comparative meta-analysis of QTL between intraspecific *Gossypium hirsutum* and interspecific *G. hirsutum* × *G. barbadense* populations. *Mol Genet Genomics* 290:1003–1025

Salmon A, Flagel L, Ying B et al (2010) Homoeologous nonreciprocal recombination in polyploid cotton. *New Phytol* 186:123–134

Sarilar V, Marmagne A, Brabant P et al (2011) *BraSto*, a *Stowaway* MITE from *Brassica*: recently active copies preferentially accumulate in the gene space. *Plant Mol Biol* 77:59–75

Sarilar V, Palacios PM, Rousselet A et al (2013) Allopolyploidy has a moderate impact on restructuring at three contrasting transposable element insertion sites in resynthesized *Brassica napus* allotetraploids. *New Phytol* 198:593–604

Sattler MC, Carvalho CR, Clarindo WR (2016) The polyploidy and its key role in plant breeding. *Planta* 243:281–296

Saunders JH (1961) The wild species of *Gossypium* and their evolutionary history. Oxford University Press, New York

Schnable PS, Ware D, Fulton RS et al (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science* 326:1112–1115

Schnable JC, Springer NM, Freeling M (2011) Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proc Natl Acad Sci U S A 108:4069–4074

Seelanan T, Brubaker CL, James McD. Stewart, et al (1999) Molecular systematics of Australian *Gossypium* section *Grandicalyx* (Malvaceae). *Syst Bot* 24:183–208

Senchina DS, Alvarez I, Cronn RC et al (2003) Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. *Mol Biol Evol* 20:633–643

Senerchia N, Felber F, Parisod C (2014) Contrasting evolutionary trajectories of multiple retrotransposons following independent allopolyploidy in wild wheats. *New Phytol* 202:975–985

Shaked H, Kashkush K, Ozkan H et al (2001) Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. *Plant Cell* 13:1749–1759

Shan C-M, Shangguan X-X, Zhao B et al (2014) Control of cotton fibre elongation by a homeodomain transcription factor *GhHOX3*. *Nat Commun* 5:5519

Shangguan X-X, Xu B, Yu Z-X et al (2008) Promoter of a cotton fibre MYB gene functional in trichomes of *Arabidopsis* and glandular trichomes of tobacco. *J Exp Bot* 59:3533–3542

Shepherd RL (1974) Breeding root-knot-resistant *Gossypium hirsutum* L. Using a resistant wild *G. barbadense* L.1. *Crop Sci* 14:687–691

Shi Y-H, Zhu S-W, Mao X-Z et al (2006) Transcriptome profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber cell elongation. *Plant Cell* 18:651–664

Singh B, Cheek HD, Haigler CH (2009) A synthetic auxin (NAA) suppresses secondary wall cellulose synthesis and enhances elongation in cultured cotton fiber. *Plant Cell Rep* 28:1023–1032

Skovsted A (1933) Cytological studies in cotton I. The mitosis and the meiosis in diploid and triploid Asiatic cotton. *Ann Bot* os-47:227–251

Skovsted A (1937) Cytological studies in cotton. *J Genet* 34:97–134

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

Small RL, Wendel JF (2000) Phylogeny, duplication, and intraspecific variation of *Adh* sequences in New World diploid cottons (*Gossypium* l., malvaceae). *Mol Phylogenet Evol* 16:73–84

Soltis DE, Soltis PS (1999) Polyploidy: recurrent formation and genome evolution. Trends Ecol Evol 14:348–352

Soltis PS, Soltis DE (2016) Ancient WGD events as drivers of key innovations in angiosperms. *Curr Opin Plant Biol* 30:159–165

Soltis DE, Soltis PS, Rieseberg LH (1993) Molecular data and the dynamic nature of polyploidy. *CRC Crit Rev Plant Sci* 12:243–273

Soltis DE, Soltis PS, Tate JA (2004) Advances in the study of polyploidy since plant speciation. *New Phytol*

Soltis DE, Albert VA, Leebens-Mack J et al (2009) Polyploidy and angiosperm diversification. *Am J Bot* 96:336–348

Soltis DE, Visger CJ, Blaine Marchant D, Soltis PS (2016) Polyploidy: pitfalls and paths to a paradigm. *Am J Bot* 103:1146–1166

Song Q, Chen ZJ (2015) Epigenetic and developmental regulation in plant polyploids. *Curr Opin Plant Biol* 24:101–109

Song K, Lu P, Tang K, Osborn TC (1995) Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. *Proc Natl Acad Sci U S A* 92:7719–7723

Song Q, Guan X, Chen ZJ (2015) Dynamic roles for small RNAs and DNA methylation during ovule and fiber development in allotetraploid cotton. *PLoS Genet* 11:e1005724

Song Q, Zhang T, Stelly DM, Chen ZJ (2017) Epigenomic and functional analyses reveal roles of epialleles in the loss of photoperiod sensitivity during domestication of allotetraploid cottons. *Genome Biol* 18:99

Splitstoser JC, Dillehay TD, Wouters J, Claro A (2016) Early pre-Hispanic use of indigo blue in Peru. *Sci Adv* 2:e1501623

Springer NM, Lisch D, Li Q (2016) Creating order from chaos: epigenome dynamics in plants with complex genomes. *Plant Cell* 28:314–325

Stanton MA, Stewart JM, Pervical AE, Wendel JF (1994) Morphological diversity and relationships in the A-Genome cottons, *Gossypium arboreum* and *G. herbaceum*. Crop Sci 34:519–527

Stebbins GL (1950) Variation and evolution in plants. New York, Columbia University Press. 643 pp.

Stebbins GL (1985) Polyploidy, hybridization, and the invasion of new habitats. *Ann Mo Bot Gard* 72:824–832

Stebbins GL et al (1971) Chromosomal evolution in higher plants. In: Chromosomal evolution in higher plants

Stephens SG (1944) The genetic organization of leaf-shape development in the genus *Gossypium*. *J Genet* 46:28–51

Stephens SG (1946) The genetics of corky; the New World alleles and their possible role as an interspecific isolating mechanism. *J Genet* 47:150–161

Stephens SG (1947) Cytogenetics of *Gossypium* and the problem of the origin of New World cottons. In: Demerec M (ed) Advances in genetics. Academic, New York, pp 431–442
Stephens SG (1949) The cytogenetics of speciation in *Gossypium*. I. Selective elimination of the donor parent genotype in interspecific backcrosses. *Genetics* 34:627–637
Stephens SG (1950) The internal mechanism of speciation in *Gossypium*. *Bot Rev* 16:115–149
Stephens SG (1958) Salt water tolerance of seeds of *Gossypium* species as a possible factor in seed dispersal. *Am Nat* 92:83–92
Stephens SG (1966) The potentiality for long range oceanic dispersal of cotton seeds. *Am Nat* 100:199–210
Stephens SG (1967) Evolution under domestication of the New World cotton (*Gossypium* spp.). *Cienciae Cultura* 19:118–134
Stephens SG (1974) Geographic and taxonomic distribution of anthocyanin genes in New World cottons. *J Genet* 61:128
Stewart JM (1992) A new cytoplasmic male sterile and restorer for cotton. In: Proceedings of the Beltwide cotton conference. National Cotton Council, Memphis. p 610
Stewart J (1995) Potential for crop improvement with exotic germplasm and genetic engineering. In: Proceeding of the world cotton research conference-I, Brisbane, Australia, February 14–17, Melbourne. pp 313–327
Sun Y, Wu Y, Yang C et al (2017) Segmental allotetraploidy generates extensive homoeologous expression rewiring and phenotypic diversity at the population level in rice. *Mol Ecol* 26:5451–5466
Suzuki H, Yu J, Ness SA et al (2013) RNA editing events in mitochondrial genes by ultra-deep sequencing methods: a comparison of cytoplasmic male sterile, fertile and restored genotypes in cotton. *Mol Gen Genomics* 288:445–457
Tan J, Walford S-A, Dennis ES, Llewellyn D (2016) Trichomes control flower bud shape by linking together young petals. *Nat Plants* 2:16093
Tang H, Woodhouse MR, Cheng F et al (2012) Altered patterns of fractionation and exon deletions in *Brassica rapa* support a two-step model of paleohexaploidy. *Genetics* 190:1563–1574
Tang M, Chen Z, Grover CE et al (2015) Rapid evolutionary divergence of *Gossypium barbadense* and *G. hirsutum* mitochondrial genomes. *BMC Genomics* 16:770
Tate JA, Ni Z, Scheen A-C et al (2006) Evolution and expression of homeologous loci in *Tragopogon miscellus* (Asteraceae), a recent and reciprocally formed allopolyploid. *Genetics* 173:1599–1611
Tu L-L, Zhang X-L, Liang S-G et al (2007) Genes expression analyses of sea-island cotton (*Gossypium barbadense* L.) during fiber development. *Plant Cell Rep* 26:1309–1320
Tuttle JR, Nah G, Duke MV et al (2015) Metabolomic and transcriptomic insights into how cotton fiber transitions to secondary wall synthesis, represses lignification, and prolongs elongation. *BMC Genomics* 16:477
Tyagi P, Gore MA, Bowman DT et al (2014) Genetic diversity and population structure in the US Upland cotton (*Gossypium hirsutum* L.). *Theor Appl Genet* 127:283–295
Udall JA, Long E, Ramaraj T, et al (2019) The genome sequence of *Gossypioides kirkii* illustrates a descending dysploidy in plants. *Frontiers in Plant Science* 10:1541
Van Deynze A, Stoffel K, Lee M et al (2009) Sampling nucleotide diversity in cotton. *BMC Plant Biol* 9:125
Vicient CM, Casacuberta JM (2017) Impact of transposable elements on polyploid plant genomes. *Ann Bot* 120:195–207
Vicient CM, Suoniemi A, Anamthawat-Jónsson K et al (1999) Retrotransposon *BARE-1* and its role in genome evolution in the genus *Hordeum*. *Plant Cell* 11:1769–1784
Vollesen K (1987) The native species of *Gossypium* (Malvaceae) in Africa, Arabia and Pakistan. *Kew Bull* 42:337–349
Walford S-A, Wu Y, Llewellyn DJ, Dennis ES (2011) GhMYB25-like: a key factor in early cotton fibre development. *Plant J* 65:785–797
Walker GP, Natwick ET (2006) Resistance to silverleaf whitefly, *Bemisia argentifolii* (Hem., Aleyrodidae), in *Gossypium thurberi*, a wild cotton species. *J Appl Entomol* 130:429–436
Wan Q, Guan X, Yang N et al (2016) Small interfering RNAs from bidirectional transcripts of *GhMML3_A12* regulate cotton fiber development. *New Phytol* 210:1298–1310

Wang GL, Dong JM, Paterson AH (1995) The distribution of *Gossypium hirsutum* chromatin in *G. barbadense* germ plasm: molecular analysis of introgressive plant breeding. *Theor Appl Genet* 91:1153–1161

Wang B, Brubaker CL, Burdon JJ (2004) Fusarium species and Fusarium wilt pathogens associated with native *Gossypium* populations in Australia. *Mycol Res* 108:35–44

Wang J, Wang H-Y, Zhao P-M et al (2010) Overexpression of a profilin (*GhPFN2*) promotes the progression of developmental phases in cotton fibers. *Plant Cell Physiol* 51:1276–1290

Wang F, Gong Y, Zhang C et al (2011a) Genetic effects of introgression genomic components from Sea Island cotton (*Gossypium barbadense* L.) on fiber related traits in upland cotton (*G. hirsutum* L.). *Euphytica* 181:41–53

Wang XQ, Feng CH, Lin ZX, Zhang XL (2011b) Genetic diversity of sea-island cotton (*Gossypium barbadense*) revealed by mapped SSRs. *Genet Mol Res* 10:3620–3631

Wang X, Wang H, Wang J et al (2011c) The genome of the mesopolyploid crop species *Brassica rapa*. *Nat Genet* 43:1035–1039

Wang K, Wang Z, Li F et al (2012) The draft genome of a diploid cotton *Gossypium raimondii*. *Nat Genet* 44:1098–1103

Wang M, Yuan D, Tu L et al (2015a) Long noncoding RNAs and their proposed functions in fibre development of cotton (*Gossypium* spp.). *New Phytol* 207:1181–1197

Wang M, Yuan D, Zhang X (2015b) Genome Sequencing. In: Fang DD, Percy RG (eds) Cotton. American Society of Agronomy, Inc.; Crop Science Society of America, Inc.; Soil Science Society of America, Inc., 5585 Guilford Road, Madison, WI 53711-1086 USA, pp 289–302

Wang B, Zhang M, Fu R et al (2016a) Epigenetic mechanisms of salt tolerance and heterosis in Upland cotton (*Gossypium hirsutum* L.) revealed by methylation-sensitive amplified polymorphism analysis. *Euphytica* 208:477–491

Wang X, Guo H, Wang J et al (2016b) Comparative genomic de-convolution of the cotton genome revealed a decaploid ancestor and widespread chromosomal fractionation. *New Phytol* 209:1252–1263

Wang X, Zhang H, Li Y et al (2016c) Transcriptome asymmetry in synthetic and natural allotetraploid wheats, revealed by RNA-sequencing. *New Phytol* 209:1264–1277

Wang M, Tu L, Lin M et al (2017) Asymmetric subgenome selection and cis-regulatory divergence during cotton domestication. *Nat Genet* 49:579–587

Wang K, Wendel JF, Jinping HUA (2018a) Designations for individual genomes and chromosomes in *Gossypium*. *J Cotton Res* 1:3

Wang P, Zhang J, Sun L et al (2018b) High efficient multisites genome editing in allotetraploid cotton (*Gossypium hirsutum*) using CRISPR/Cas9 system. *Plant Biotechnol J* 16:137–150

Wang M, Tu L, Yuan D et al (2019) Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*. *Nat Genet* 51:224–229

Ware JO (1951) Origin, rise and development of American Upland cotton varieties and their status at present. University of Arkansas College of Agriculture, Agricultural Experiment Station, Fayetteville

Watt SG (1907) The wild and cultivated cotton plants of the world: a revision of the genus *Gossypium*, framed primarily with the object of aiding planters and investigators who may contemplate the systematic improvement of the cotton staple. Longmans, Green, and Co.

Wayne Smith C, Tom Cothren J (1999) Cotton: origin, history, technology, and production. Wiley, New York

Wei Y, Xu Y, Lu P et al (2017) Salt stress responsiveness of a wild cotton species (*Gossypium klotzschianum*) based on transcriptomic analysis. *PLoS One* 12:e0178313

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

Wendel JF (2015) The wondrous cycles of polyploidy in plants. *Am J Bot* 102:1753–1756

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

Wendel JF, Cronn RC (2003) Polyploidy and the evolutionary history of cotton. *Adv Agron* 78:139–186

Wendel JF, Grover CE (2015) Taxonomy and evolution of the cotton genus, *Gossypium*. In: Agronomy monograph

Wendel JF, Percival AE (1990) Molecular divergence in the Galapagos Islands—Baja, California species pair, *Gossypium klotzschianum* and *G. davidsonii* (Malvaceae). *Plant Syst Evol* 171:99–115

Wendel JF, Olson PD, Stewart JM (1989) Genetic diversity, introgression, and independent domestication of Old World cultivated cottons. *Am J Bot* 76:1795–1806

Wendel JF, Stewart JM, Rettig JH (1991) Molecular evidence for homoploid reticulate evolution among Australian species of *Gossypium*. *Evolution* 45:694–711

Wendel JF, Brubaker CL, Percival AE (1992) Genetic diversity in *Gossypium hirsutum* and the origin of Upland cotton. *Am J Bot* 79:1291–1310

Wendel JF, Rowley R, Stewart JM (1994) Genetic diversity in and phylogenetic relationships of the Brazilian endemic cotton, *Gossypium mustelinum* (Malvaceae). *Plant Syst Evol* 192:49–59

Wendel JF, Schnabel A, Seelanan T (1995a) An unusual ribosomal DNA sequence from *Gossypium gossypioides* reveals ancient, cryptic, intergenomic introgression. *Mol Phylogenet Evol* 4:298–313

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

Wendel JF, Cronn RC, Alvarez I et al (2002) Intron size and genome size in plants. *Mol Biol Evol* 19:2346–2352

Wendel JF, Brubaker C, Alvarez I et al (2009) Evolution and natural history of the cotton genus. In: Paterson AH (ed) Genetics and genomics of cotton. Springer US, New York, pp 3–22

Wendel JF, Brubaker CL, Seelanan T (2010) The origin and evolution of *Gossypium*. In: Physiology of cotton, pp 1–18

Wendel JF, Flagel LE, Adams KL (2012) Jeans, genes, and genomes: cotton as a model for studying polyploidy. In: Polyploidy and genome evolution

Westengen OT, Huamán Z, Heun M (2005) Genetic diversity and geographic pattern in early South American cotton domestication. *Theor Appl Genet* 110:392–402

Wittkopp PJ, Haerum BK, Clark AG (2004) Evolutionary changes in cis and trans gene regulation. *Nature* 430:85–88

Woodhouse MR, Schnable JC, Pedersen BS et al (2010) Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homeologs. *PLoS Biol* 8:e1000409

Woodhouse MR, Cheng F, Pires JC et al (2014) Origin, inheritance, and gene regulatory consequences of genome dominance in polyploids. *Proc Natl Acad Sci U S A* 111:5283–5288

Wu H, Tian Y, Wan Q et al (2018) Genetics and evolution of MIXTA genes regulating cotton lint fiber development. *New Phytol* 217:883–895

Xu Q, Xiong G, Li P et al (2012) Analysis of complete nucleotide sequences of 12 *Gossypium* chloroplast genomes: origin and evolution of allotetraploids. *PLoS One* 7:e37128

Xu C, Bai Y, Lin X et al (2014) Genome-wide disruption of gene expression in allopolyploids but not hybrids of rice subspecies. *Mol Biol Evol* 31:1066–1076

Yang SS, Samuel Yang S, Cheung F et al (2006) Accumulation of genome-specific transcripts, transcription factors and phytohormonal regulators during early stages of fiber cell development in allotetraploid cotton. *Plant J* 47:761–775

Yik CP, Birchfield W (1984) Resistant germplasm in *Gossypium* species and related plants to *Rotylenchulus reniformis*. *J Nematol* 16:146–153

Yoo M-J, Wendel JF (2014) Comparative evolutionary and developmental dynamics of the cotton (*Gossypium hirsutum*) fiber transcriptome. *PLoS Genetics* 10:e1004073

Yoo M-J, Szadkowski E, Wendel JF (2013) Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity* 110:171–180

Yoo M-J, Liu X, Pires JC et al (2014) Nonadditive gene expression in polyploids. *Annu Rev Genet* 48:485–517

Yuan D, Tang Z, Wang M et al (2015) The genome sequence of Sea-Island cotton (*Gossypium barbadense*) provides insights into the allopolyploidization and development of superior spinnable fibres. *Sci Rep* 5:17662

Yu-xiang W, Jin-hong C, Qiu-ling H, Shui-jin Z (2013) Parental origin and genomic evolution of tetraploid *Gossypium* species by molecular marker and GISH analyses. *Caryologia* 66:368–374
Zafar SA, Noor MA, Waqas MA et al (2018) Temperature extremes in cotton production and mitigation strategies. In: Mehboob-Ur-Rahman, Zafar Y (eds) Past, present and future trends in cotton breeding. InTech
Zhang T, Chen T (2012) Cotton pistil drip transformation method. *Methods Mol Biol* 847:237–243
Zhang T, Endrizzi JE (2015) Cytology and cytogenetics. In: Cotton. American Society of Agronomy, Inc., Crop Science Society of America, Inc., and Soil Science Society of America, Inc., Madison, pp 129–154
Zhang JF, Stewart JM (1999) Cytoplasmic male sterility based on *Gossypium sturtianum* cytoplasm (CMS-C1): Characterization and genetics of restoration. *Proc Cotton Res Meeting, Ark Agric Exp Stn Special Rep* 193:269–272
Zhang JF, Stewart JM (2001) Inheritance and genetic relationships of the D8 and D2-2 restorer genes for cotton cytoplasmic male sterility. *Crop Sci* 41:289–294
Zhang J, Lu Y, Adragna H, Hughs SE (2005) Genetic improvement of New Mexico Acala cotton germplasm and their genetic diversity. *Crop Sci* 45:2363–2373
Zhang J, Fang H, Zhou H et al (2013) Inheritance and transfer of thrips resistance from Pima cotton to Upland cotton. *J Cotton Sci* 17:163–169
Zhang J, Percy RG, McCarty JC (2014) Introgression genetics and breeding between Upland and Pima cotton: a review. *Euphytica* 198:1–12
Zhang T, Hu Y, Jiang W et al (2015) Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat Biotechnol* 33:531–537
Zhang F, Zhu G, Du L et al (2016) Genetic regulation of salt stress tolerance revealed by RNA-Seq in cotton diploid wild species, *Gossypium davidsonii*. *Sci Rep* 6:20582
Zhang H, Mittal N, Leamy LJ et al (2017a) Back into the wild-Apply untapped genetic diversity of wild relatives for crop improvement. *Evol Appl* 10:5–24
Zhang X, Hu D-P, Li Y et al (2017b) Developmental and hormonal regulation of fiber quality in two natural-colored cotton cultivars. *J Integr Agric* 16:1720–1729
Zhang Z, Ge X, Luo X et al (2018) Simultaneous editing of two copies of *Gh14-3-3d* confers enhanced transgene-clean plant defense against *Verticillium dahliae* in allotetraploid upland cotton. *Front Plant Sci* 9:842
Zhao X-P, Si Y, Hanson RE et al (1998) Dispersed repetitive DNA has spread to new genomes since polyploid formation in cotton. *Genome Res* 8:479–492
Zhao Y-L, Yu S-X, Ye W-W et al (2010) Study on DNA cytosine methylation of cotton (*Gossypium hirsutum* L.) genome and its implication for salt tolerance. *Agric Sci China* 9:783–791
Zhao F, Fang W, Xie D et al (2012) Proteomic identification of differentially expressed proteins in *Gossypium thurberi* inoculated with cotton *Verticillium dahliae*. *Plant Sci* 185-186:176–184
Zhao Y, Wang H, Chen W et al (2015) Genetic diversity and population structure of elite cotton (*Gossypium hirsutum* L.) germplasm revealed by SSR markers. *Plant Syst Evol* 301:327–336
Zhao X, Meng Z, Wang Y et al (2017) Pollen magnetofection for genetic modification with magnetic nanoparticles as gene carriers. *Nat Plants* 3:956–964
Zhao B, Cao J-F, Hu G-J et al (2018) Core cis-element variation confers subgenome-biased expression of a transcription factor that functions in cotton fiber elongation. *New Phytol* 218:1061–1075
Zheng D, Ye W, Song Q et al (2016) Histone modifications define expression bias of homoeologous genomes in allotetraploid cotton. *Plant Physiol* 172:1760–1771
Zhu SJ, Reddy N, Jiang YR (2005) Introgression of a gene for delayed pigment gland morphogenesis from *Gossypium bickii* into upland cotton. *Plant Breed* 124:590–594