GENETICS

Cotton genome evolution and features of its structural and functional organization

Ksenia Strygina, Elena Khlestkina, and Larisa Podolnaya

Federal Research Center N. I. Vavilov All-Russian Institute of Plant Genetic Resources, Bol'shaya Morskaya ul., 42–44, Saint Petersburg, 190000, Russian Federation

Address correspondence and requests for materials to Ksenia Strygina, k.strygina@vir.nw.ru

Abstract

Allotetraploid cotton *Gossypium hirsutum* L. is not only an important crop, but also a model organism used to study such processes as polyploidization, plant genome evolution and the influence of polyploidy on gene expression. The present article provides a review of studies devoted to the taxonomy of the genus *Gossypium*, the evolution of the genomes of its representatives (including 45 diploid and 7 allotetraploid species), and the functional divergence of duplicated copies of the same genes in allotetraploid species. The discussion concerns the areas of individual species' origin, as well as the reasons of the high variation in genome size (from ~880 Mb to ~2400 Mb), which was influenced by both full-genome duplications and the spread of mobile genetic elements. The data support the fact that the expression of genes in allotetraploid cotton changes as a result of polyploidization, and that one of the two subgenomes dominates in the formation of one or another trait. The considered data shed light on the features of the evolution of plant genes and genomes.

Keywords: Allopolyploid genome, cotton, evolution, functional divergence, gene duplication, *Gossypium*, homoeologous genes, mobile genetic elements.

1. Genus *Gossypium* L.

Cotton (genus *Gossypium* L.) belongs to the tribe Gossypieae Alef., which comprises nine genera (Fryxell, 1968, 1978; Phuphathanaphong, 2006). Five genera include a small number of species; they are monotypic and are characterized by narrow distribution areas: *Cephalohibiscus* Ulbr. (one species; New Guinea, Solomon Islands), *Gossypioides* Skovst. ex J. B. Hutch. (two species; East Africa, Madagascar), *Kokia* Lewton (six species; Hawaii), *Lebronnecia* Fosberg (one species; Marquesas Islands) and *Thepparatia* Phuph. (one species; northern Thailand). The other four genera are larger and represent groups with moderately wide geographical ranges: *Cienfuegosia* Cav. (25 species; neotropical realm, part of Africa), *Hampea* Schltdl. (21 species; neotropical realm), *Thespesia* Sol. ex Corrêa (17 species; tropics) and the largest and most widespread genus *Gossypium* L. (52 species). These representatives grow in the tropical and subtropical regions of the Old and New Worlds (Fig. 1) (Fryxell, Craven and Stewart, 1992; Wendel et al., 2009).

Despite wide distribution and morphological diversity, the genus *Gossypium* represents a single phylogenetic group (Fryxell et al., 1992; Wendel and Grover, 2015). The closest relatives of cotton are the Hawaiian endemic genus *Kokia* and the Afro-Madagascan genus *Gossypioides* (Seelanan, Schnabel and Wendel, 1997). Their divergence occurred during the Miocene 10–15 million years ago (MYA) with the subsequent spread of *Gossypium* over almost all continents (Fig. 2) (Wendel et al., 2009; Wendel and Grover, 2015).

As the genus *Gossypium* was undergoing its formation (5–10 MYA) and spreading to various environments, the genome of *Gossypium* was undergoing significant changes and rearrangements (Fig. 2) (Hendrix and Stewart, 2005). It is reflected in such phenotypic features as the type of ontogenesis, plant life-form,

Citation: Strygina, K., Khlestkina, E., and Podolnaya, L. 2020. Cotton genome evolution and features of its structural and functional organization. Bio. Comm. 65(1): 15–27. https://doi.org/10.21638/spbu03.2020.102

Author's information: Ksenia Strygina, PhD, Researcher, orcid.org/0000-0001-6938-1348; Elena Khlestkina, Dr. of Sci. in Biology, Professor RAS, Head of Laboratory, orcid.org/0000-0002-8470-8254; Larisa Podolnaya, PhD, Leading Researcher, orcid.org/0000-0002-4962-1989

Manuscript Editor: Alla Krasikova, Department of Cytology and Histology, Faculty of Biology, Saint Petersburg State University, Saint Petersburg, Russia

Received: September 2, 2019;

Revised: December 16, 2019;

Accepted: January 12, 2020;

Funding: The present review was carried out within VIR project No. 0481-2019-0001.

Competing interests: The authors have declared that no competing interests exist.

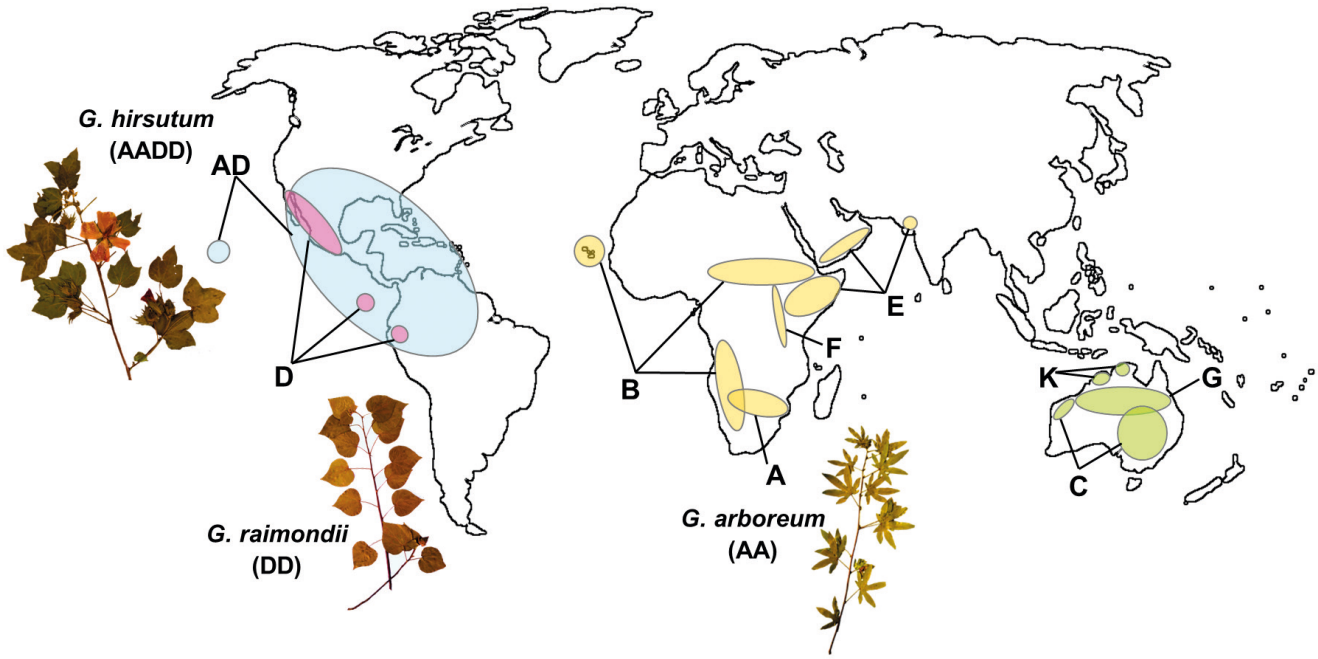

Fig. 1. Distribution of the genus *Gossypium*. Yellow indicates the distribution of the *Gossypium* subgenus representatives (A, B, E, F genomes); green — subgenus *Sturtia* representatives (C, G, K genomes); purple — subgenus *Houzingenia* representatives (D genome); blue — subgenus *Karpas* representatives (AD genome). Photos of *G. arboreum* (AA), *G. raimondii* (DD) and *G. hirsutum* (AADD) cotton samples from the herbarium collection of the N. I. Vavilov All-Russian Institute of Plant Genetic Resources.

corolla color, leaf shape, seed shape, distribution mode, etc. According to the generally accepted system by P. A. Fryxell (1992), supplemented by newly described species, the genus *Gossypium* includes 45 diploid (2n = 2x = 26) and 7 allotetraploid (2n = 4x = 52) species (Tab. 1) (Fryxell, 1992; Fryxell et al., 1992; Grover et al., 2014; Yu et al., 2014; Gallagher et al., 2017). The main centers of genus diversity are the arid areas of Australia, Africa, Arabia, the Indian subcontinent, the Galapagos and Hawaiian Islands, and Central and South Americas (Fig. 1) (Fryxell et al., 1992; Wendel et al., 2009).

Based on the analysis of morphological features, the nature of distribution areas, cytogenetic and molecular genetic data and the relative fertility of interspecific hybrids, cotton species fall into eight genomic groups (A–G and K) united into four subgenera; one of these groups is represented by AD allotetraploid representatives (Tab. 1) (Webber, 1938; Endrizzi, Turcotte and Kohel, 1985; Fryxell, 1992; Wendel and Grover, 2015). The three subgenera that include diploid species have separate distribution areas, namely Africa/Arabia (A, B, E and F genomes), Australia (C, G and K genomes) and America (D genome) (Fig. 1, Tab. 1) (Wendel and Cronn, 2003). The monophyletic origin of the allotetraploid group (AD genome) in the regions of the New World is associated with the spread of the A genome donor, related to *G. arboreum* L. or *G. herbaceum* L., from Africa or Asia over long distances to the New World, and subsequent hybridization (about 1–1.5 MYA) with the American representative of the D genomic group, genetically close to *G. raimondii* Ulbr. (Fig. 1) (Skovsted, 1933, 1934; Lemeshev, 1991; Wendel and Albert, 1992; Wendel and Cronn, 2003; Paterson et al., 2012; Lu et al., 2018).

1.1. Diploid species of the genus *Gossypium*

Subgenus Gossypium Tod. (A, B, E, F). According to the last taxonomic interpretation, the subgenus *Gossypium* Tod. includes 14 species from Africa and Arabia (Fryxell, 1992). These species exhibit significant cytogenetic diversity that corresponds to the A, B, E, and F genomic groups (Tab. 1, Fig. 2) (Seelanan et al., 1997; Wendel and Albert, 1992). It is supposed that such a variation in genomes compared to the relative uniformity of cotton found in the New World (see below) may indicate the African origin of the genus *Gossypium* (Wendel and Grover, 2015).

To date, the most well-studied is the A genomic group, which includes two cultivated cotton species — *G. herbaceum* L. and *G. arboreum* L. (Wendel, Olson and Stewart, 1989). The B genomic group is represented by four African species, while the F genomic group is represented by only one species — *G. longicalyx* Hutch. and Lee (Tab. 1). Its cytogenetic and morphological differences from other *Gossypium* members are probably explained by its geographical isolation (Fryxell, 1971; Vollesen, 1987; Wendel and Grover, 2015).

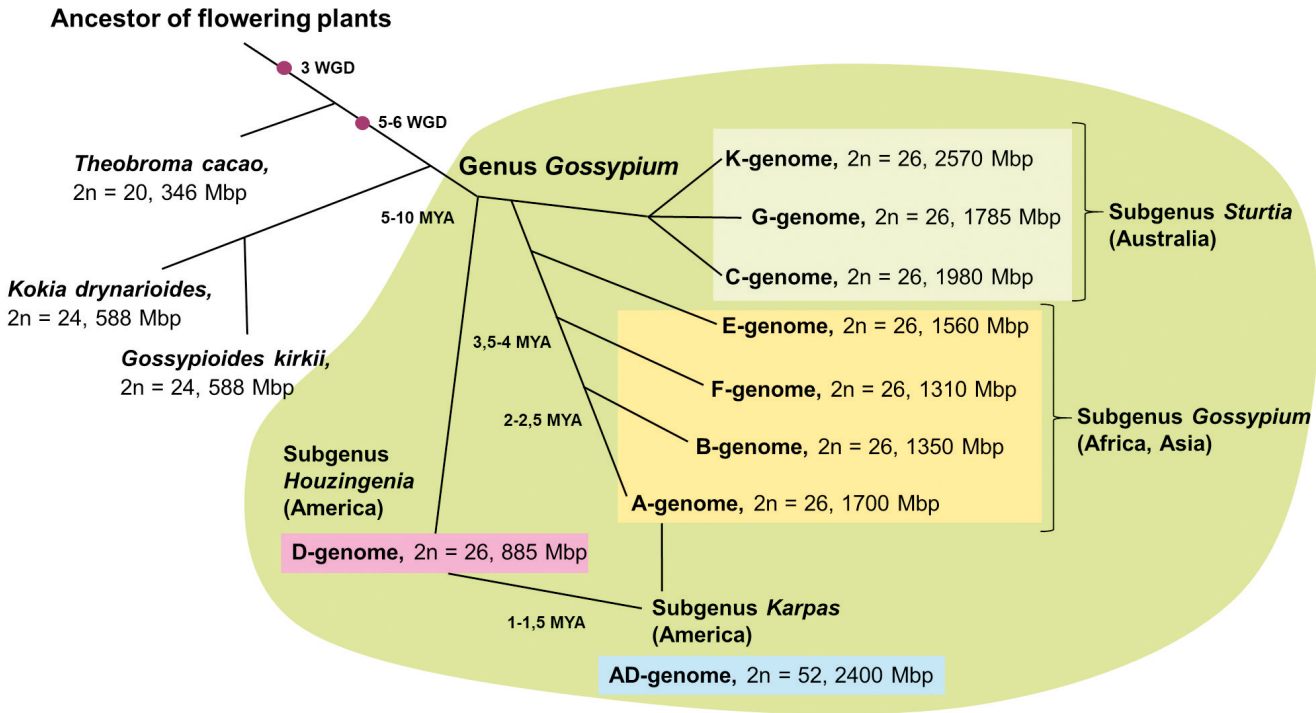

Fig. 2. A modern look at the phylogeny of the genus *Gossypium*. Genome sizes, number of whole-genome duplications (WGD) and genome divergence time (MYA — million years ago) are based on the works by Wendel and Cronn, 2003; Hendrix and Stewart, 2005; Wendel and Grover, 2015; Lu et al., 2018.

The E genome encompasses three well-studied species and five poorly studied species (the so-called *G. somalense* complex) (Tab. 1) (Vollesen, 1987; Golubets, 1991). Of these species, only *G. somalense* has been studied cytogenetically; other species of the group have been described from several herbarium specimens, and therefore their individuality is doubtful (Tab. 1). The problem is that the distribution area of these species is located in the Horn of Africa, where it is not possible to organize collecting missions.

Regarding *G. trifurcatum* Vollesen, which also grows in Somalia and has been described only from an herbarium specimen, its taxonomic status is unclear. Fryxell (1992) separated this species into an individual *Serrata* section. This species may belong to both B and E genomes (Tab. 1) (Wendel and Grover, 2015).

Subgenus Sturtia (R. Br.) Tod. (C, G, K). The Australian subgenus *Sturtia* (R. Br.) Tod. includes three genomic groups (C, G and K) represented by 4, 3 and 11 diploid cotton species, respectively (Fig. 2, Tab. 1) (Fryxell, 1978; Wendel et al., 2009; Tiwari, Zhang and Stewart, 2014; Wendel and Grover, 2015). The taxonomy of species in this group still needs to be refined (Podolnaya, 1991). Nevertheless, their division into three genomic groups according to DNA sequence data (Seelanan et al., 1997) is consistent with their formal separation into the *Sturtia* (C genome), *Hibiscoidea* (G genome) and *Grandicalyx* (K genome) taxonomic sections (Fryxell, 1978; Tiwari et al., 2014).

Most species of the subgenus *Sturtia* are sympatric in their areas of distribution and occur exclusively in the northern region of Australia (Seelanan et al., 1999; Wendel, Stewart and Rettig, 1991). The species from the C and G genomic groups are available in many genetic resources collections around the world and have been thoroughly studied. Despite this, the tendency for hybridization and genetic material introgression in the C genome species poses certain complications (Seelanan et al., 1999; Cronn and Wendel, 2003; Tiwari et al., 2014).

The taxonomic status of the K genome representatives is unclear due to insufficient knowledge (Fryxell, 1978; Seelanan et al., 1997, 1999; Tiwari et al., 2014). In terms of the nature of the distribution area, as well as many morphological characters, including the seedlings structure, life-form and peculiarities of seed distribution, the species of the *Grandicalyx* section differ sharply from other representatives of the subgenus *Sturtia*. On this basis, suggestions are made to change the rank of the taxon (Podolnaya, 1991). However, representatives of this section are almost absent in collections, and it complicates the research (Campbell et al., 2010).

Subgenus Houzingenia Fryx. (D). The subgenus *Houzingenia* Fryx. is the best-studied group represented by thirteen D genomic diploids from the New World (Fig. 2, Tab. 1) (Seelanan et al., 1997; Ulloa, 2014; Wendel and Grover, 2015 Grover et al., 2018). These species have unusually small genomes for the genus *Gossypium* (Fig. 2) (Hendrix and Stewart, 2005). Not a single rep-

Table 1. Taxonomy of the genus *Gossypium* in accordance with CottonGen database (Yu et al., 2014). Cultivated cotton species are bolded in the table

Subgenus	Genomic group	Species	Genome	Geographical distribution
Gossypium Tod.	A	***G. herbaceum* L.**	A_1	Africa, Asia
		***G. arboreum* L.**	A_2	Asia
	B	*G. anomalum* Wawr. and Peyer	B_1	Africa
		G. triphyllum (Harv. and Sand.) Hochr.	B_2	Africa
		G. capitis-viridis (Harv. and Sand.) Hochr.	B_3	Cape Verde Islands
		G. trifurcatum Vollesen	B^*	Africa
	E	*G. stocksii* Mast. ex. Hook.	E_1	Arabia
		G. somalense (Gurke) Hutch.	E_2	Arabia
		G. areysianum (Defl.) Hutch.	E_3	Arabia
		G. incanum (Schwartz) Hillc.	E_4	Arabia
		G. benadirense Mattei	E^*	Arabia
		G. bricchettii (Ulbr.) Vollesen	E^*	Arabia
		G. vollesenii Fryx.	E^*	Arabia
	F	*G. longicalyx* Hutch. and Lee	F_1	Arabia
Sturtia (R. Br.) Tod.	C	*G. sturtianum* J. H. Willis	C_1	Australia
		G. nandewarense (Derera) Fryx.	$C_{(1-n)}$	Australia
		G. robinsonii F. Muell.	C_2	Australia
		G. pilosum Fryxell	C_{10}	Australia
	G	*G. bickii* Prokh.	G_1	Australia
		G. australe F. Muell.	G_2	Australia
		G. nelsonii Fryxell	G_3	Australia
	K	*G. costulatum* Tod.	K_1	Australia
		G. populifolium (Benth.) Tod.	K_2	Australia
		G. cunninghamii Tod.	K_3	Australia
		G. pulchellum (C. A. Gardner) Fryxell	K_4	Australia
		G. anapoides J. M. Stewart, Craven, Brubaker and Wendel	K_6	Australia
		G. enthyle Fryxell et al.	K_7	Australia
		G. exiguum Fryxell et al.	K_8	Australia
		G. londonderriense Fryxell et al.	K_9	Australia
		G. marchantii Fryxell et al.	K_{10}	Australia
		G. nobile Fryxell et al.	K_{11}	Australia
		G. rotundifolium Fryxell et al.	K_{12}	Australia
Houzingenia Fryx.	D	*G. thurberi* Tod.	D_1	Mexico and Southwestern USA
		G. armourianum Kearney	$D_{(2-1)}$	Mexico
		G. harknessii Brandegee	$D_{(2-2)}$	Mexico
		G. davidsonii Kellogg	$D_{(3-d)}$	Mexico
		G. klotzschianum Andersson	$D_{(3-k)}$	Galapagos Islands
		G. aridum (Rose and Standl.) Skovst.	D_4	Mexico
		G. raimondii Ulbr.	D_5	Peru
		G. gossypioides (Ulbr.) Standl.	D_6	Mexico
		G. lobatum Gentry	D_7	Mexico
		G. trilobum (Moc. and Sess. ex DC.) Skov.	D_8	Mexico
		G. laxum L. Ll. Phillips	D_9	Mexico
		G. turneri Fryxell	D_{10}	Mexico
		G. schwendimanii Fryxell and S. D. Koch	D_{11}	Mexico
Karpas Raf.	AD	***G. hirsutum* L.**	$(AD)_1$	Southern Mexico
		***G. barbadense* L.**	$(AD)_2$	Northwestern Southern America
		G. tomentosum Nutt. ex Seem.	$(AD)_3$	Hawaii
		G. mustelinum Miers ex G. Watt	$(AD)_4$	Brazil
		G. darwinii G. Watt	$(AD)_5$	Galapagos Islands
		G. ekmanianum Wittmack	$(AD)_6$	Dominican Republic
		G. stephensii J. Gallagher et al.	$(AD)_7$	Wake Atoll, Pacific Ocean

resentative of this group is capable of producing textile fiber.

The origin of the D genomic group is associated with the spread of a *Gossypium* ancestral form from Africa over long distances about 5–10 MYA. Probably, it spread to western Mexico — the center of D genomic diversity (Wendel and Grover, 2015). The appearance of endemic species in Peru (*G. raimondii*) and the Galapagos Islands (*G. klotzschianum*) is associated with the later spread of the D genome ancestor to these territories, probably during the Pleistocene (Wendel and Percival, 1990).

1.2. Polyploid species of the genus *Gossypium*

Subgenus Karpas Raf. (AD). The American polyploid species of cotton represent a monophyletic allotetraploid group containing two genomes — the A genome from Africa or Asia, close to the existing *G. herbaceum* and *G. arboreum*, and the D genome, close to *G. raimondii* (Fig. 1, 2) (Endrizzi et al., 1985; Small and Wendel, 1999; Paterson et al., 2012). By now, seven species have been described (Tab. 1). Two of them, *G. ekmanianum* and *G. stephensii*, have been recently discovered as endemics of the Dominican Republic and Wake Atoll in the Pacific, respectively (Grover et al., 2014; Gallagher et al., 2017). Two other species, *G. hirsutum* and *G. barbadense*, are cultivated species of cotton and are of great commercial importance today (Fang, 2018).

Thus, representatives of the genus *Gossypium*, including 45 diploids and 7 allotetraploids, fall into four subgenera: *Gossypium* (A, B, E, F genomes), *Sturtia* (C, G, K genomes), *Houzingenia* (D genome) and *Karpas* (AD genomes). The status of individual taxa in the genus *Gossypium* is still unclear, especially within the subgenera *Gossypium* and *Sturtia*. The study of representatives of these taxa is very difficult due to the inaccessibility of their growing areas, the poor representation of representatives in collections and the tendency to produce interspecific hybrids. This indicates the temporary nature of the majority of the taxonomy of *Gossypium* species.

2. The evolution of the cotton genome

The morphology of chromosomes is similar among the closely related species, and it is reflected in their ability to form interspecific hybrids that exhibit normal meiotic pairing of chromosomes and high fertility of F1 hybrids. The species of each of the genomic groups of the genus *Gossypium* have the same basic chromosome number (n = 13), however, the DNA content in each genome varies significantly — from ~880 Mb (D genome; 2C = 1.81 pg) to ~2400 Mb (K genome; 2C = 5.26 pg) (Fig. 2) (Hendrix and Stewart, 2005). It is believed that such a change in DNA content has been caused by the modification of repetitive DNA sequences (Geever, Katterman and Endrizzi, 1989). Along with that, the variation of the amount of DNA in diploid species offers a good model system for studying the causes of the genome size variation.

2.1. Polyploidization

Polyploidization is an important process in plant speciation. It underlies the ample diversity of angiosperms (Alix, Gérard, Schwarzacher and Heslop-Harrison, 2017). The assumption that diploid cotton is a paleopolyploid organism was first made about 90 years ago when studying the behavior of chromosomes in the metaphase of meiosis (Denham, 1924; Lawrence, 1931; Davie, 1933; Skovsted, 1933). It was later shown that multiple duplicated segments of chromosomes found in the genomes of diploid species of cotton demonstrate that the ancestor of *Gossypium* underwent ancient polyploidization cycles with subsequent genome rearrangements and diploidization (Brubaker, Paterson and Wendel, 1999; Cronn, Zhao, Paterson and Wendel, 1996; Paterson, 2009; Paterson et al., 2012; Jiao and Paterson, 2014; Renny-Byfield et al., 2014; Renny-Byfield, Gong, Gallagher and Wendel, 2015; Rong et al., 2010). In addition to three acts of genome duplication that occurred in the ancestor of all flowering plants, the diploid ancestor of cotton underwent an additional five to six duplications shortly after divergence from the ancestor of *Theobroma cacao* L. about 60 MYA (Fig. 2) (Bowers, Chapman, Rong and Paterson, 2003; Paterson et al., 2012; Renny-Byfield et al., 2015). Thus, modern cottons are at least paleooctaploids.

2.2. Mobile genetic elements

An increase in the number of mobile genetic elements (MGEs) along with polyploidization is probably one of the main factors determining the size of the plant genome (San Miguel and Bennetzen, 1998; Zhao et al., 1998; Bennetzen, Ma and Devos, 2005; Hawkins et al., 2006; Ozkan et al., 2010). The comparative analysis of *Gossypium* genomes with *T. cacao* and *Arabidopsis thaliana* (L.) Heynh. showed that the genomes of *Gossypium* species contain a higher number of MGEs (Wu et al., 2017). This may indicate that in addition to the full-genome duplication, a change in the genome size in the genus *Gossypium* is associated with the abundance of MGEs, in particular, Long Terminal Repeat (LTR) retrotransposons (Wu et al., 2017).

Based on the comparative analysis of the nucleotide sequences in the genomes of diploid species *G. raimondii* (D_5 genome, 885 Mb) and *G. arboreum* (A_2 genome, 1746 Mb), as well as of the allotetraploid species *G. hirsutum* ($(AD)_1$ genome, 2173 Mb) (Paterson et al., 2012; Wang et al., 2012; Page et al., 2013; Kim, 2015; Li et al., 2015), it was found that the fraction of MGEs in the genomes of *G. arboreum*, *G. raimondii* and *G. hirsutum* is 57.09 %, 67.64 % and 67.36 %, respectively (Dillehay, Rossen, Andres and Williams, 2007; Wang et al., 2012). Moreover, A and AD genomes carry a significantly larger number of LTR retrotransposons than the D genome (Tab. 2). It was shown that the LTR-Copia sequences had been accumulating at a higher rate in the *Gossypium* species with the smallest genome (*G. raimondii*), while the LTR-Gypsy sequences are common in the species with larger genomes (Hawkins et al., 2006; Page et al., 2013). At the same time, it was found that one LTR-

Table 2. MGEs content in genomes of *G. arboreum* (A_2 genome), *G. raimondii* (D_5 genome) and *G. hirsutum* ($(AD)_1$ genome) (according to Wang et al., 2015)

MGE	*G. arboreum*, % of genome	*G. raimondii*, % of genome	*G. hirsutum*, % of genome
LINE	1.20	1.50	1.56
SINE	0.01	0.09	0.03
LTR-Gypsy	55.80	33.80	52.54
LTR-Copia	5.50	11.10	8.36
Others	5.13	10.60	4.87
Total	67.64	57.09	67.36

Gypsy group, GORGE3 (Gossypium retrotransposable gypsy-like element), had undergone mass distribution in large cotton genomes to become the main reason for their size change (Hawkins et al., 2006).

MGEs and the fiber-forming ability. The most valuable feature of a number of species in the *Gossypium* genus is the ability to form unicellular fibers (trichomes) of different size on the seed surface (Kim, 2015). In cultivated species, these fibers are used for spinning. Among the different types of cotton capable of forming fiber, there are significant differences in its properties, since the two genomes, A and D, in the genus *Gossypium* make an unequal contribution to the development of fiber (see below) (Paterson et al., 2012; Xu et al., 2015). Thus, the allotetraploid *G. hirsutum* produces fibers longer than 3 cm, and the diploid *G. arboreum* produces 1.3–1.5 cm long fibers (Li et al., 2015). However, there are more fiber quality-related sites in the D subgenome in *G. hirsutum* than in the A subgenome, despite the fact that the relative of the D genome progenitor *G. raimondii* does not produce spinning fiber (Jiang, Wright, El-Zik and Paterson, 1998).

Over the past two decades, many genes that are involved in the regulation of growth and development of cotton fibers have been revealed (Shi et al., 2006; Wu et al., 2006; Taliercio and Boykin, 2007; Wu et al., 2007; Wang et al., 2010; Zhang et al., 2010; Walford, Wu, Llewellyn and Dennis, 2011; Kim, 2015). Along with that, it turned out that a large number of MGEs in *Gossypium* genomes are located close (within 5 kbp) to the fiber development genes, which allows supposing that these genetic elements could contribute to this process (Wang et al., 2012; Li et al., 2014, 2015; Kun Wang, Huang and Zhu, 2016; Wu et al., 2017). For example, the promoter region of the gene encoding GhMYB25 transcription factor, which is necessary for fiber development, has shown the LTR-Copia retrotransposon insertion (3928 bp) only in the D subgenome (Fig. 3). That positively correlates with a higher expression of the D genome homoeolog in *G. hirsutum* (Zhang et al., 2010; Walford et al., 2011; Wang et al., 2016). A similar mutation was noted for the *ethylene response factor* (*ERF*) gene involved in the development of trichomes: the LINE retrotransposon insertion into the *GhERF* promoter in the D subgenome causes an increase in the expression level of this homoeolog compared to its A genomic copy (Shi et al., 2006; Qin et al., 2007; Wang et al., 2016).

A new look at the evolution of the genus Gossypium. It was previously believed that diploid forms of cotton carrying the A genome appeared less than 5 MYA after the divergence from the ancestor of the F genome forms (Wendel and Albert, 1992). Allotetraploid species were thought to have formed as a result of interspecific hybridization about 1–2 MYA (Wendel and Albert, 1992).

In 2018, a new family of LTR elements named CICR (Chinese Institute of Cotton Research) was identified in the genus *Gossypium* (Lu et al., 2018). It was shown that these MGEs are widespread in all chromosomes of the A and B genomes, but are almost absent in the genomes C–G (Cui et al., 2016; Lu et al., 2018). The analysis of CICR showed that the A and D genomes diverged at least 4 MYA (before the appearance of CICR), which coincides with the results of previous studies on the divergence time of the ancestors of these genomes about 5–10 MYA (Fig. 2) (Wendel and Albert, 1992; Senchina et al., 2003; Liu et al., 2015; Zhang et al., 2015). The divergence of the ancestors of the C–G genomes occurred probably about 3.5–4 MYA, i.e., approximately during the appearance of CICR elements (Lu et al., 2018). Besides, according to the distribution of these mobile elements, the A and B genomes are the closest to each other among the genomes of the genus *Gossypium*, having diverged about 2.5 MYA (Lu et al., 2018). It contradicts the previous information that the F genome is more similar to the A than to the B genome (Fig. 2) (Grover et al., 2004).

The formation of allotetraploid cotton about 1–1.5 MYA as a result of hybridization between the A and D genome ancestors coincides with the results of previous studies (Wendel, 1989; Wendel and Albert, 1992; Wendel and Cronn, 2003; Senchina et al., 2003; Li

et al., 2015; Zhang et al., 2015). However, it is assumed that allotetraploid cotton was formed after the CICR family silencing, since these MGEs were preserved in the A subgenome, though not transferred to the D subgenome (Lu et al., 2018).

Thus, the insertions of MGEs and their polymorphism among genomes and subgenomes could be a key factor in the evolution of cotton, as well as in the process of artificial selection of traits that determine fiber properties.

3. Expression of homoeologous genes in cotton

The key point in the evolution of genomes of polyploid organisms relates to the regulation of intergenomic interactions (including the nuclear-cytoplasmic ones), on the one hand, and normalization of the consequences of the gene duplication, on the other hand (Panchy, Lehti-Shiu and Shiu, 2016; Sattler, Carvalho and Clarindo, 2016).

The fusion of the A and D genomes of the allotetraploid cotton ancestors into the genome of one organism with the A genome cytoplasm caused a change in the level and pattern of expression of genes from both genomes due to new interactions. As a consequence, the expression of some homoeologous genes underwent significant changes due to the merging of regulating factors and their target genes (Riddle and Birchler, 2003; Birchler, Riddle, Auger and Veitia, 2005; Chen, 2007; Panchy et al., 2016; Sattler et al., 2016). On the other hand, the suppression of some homoeologous genes expression occurred as a compensation of the change in gene dosage that accompanied polyploidy (Osborn et al., 2003; Birchler et al., 2005; Sattler et al., 2016).

3.1. Changes in the homoeologous genes expression

The first evidence that polyploidy within the genus *Gossypium* is accompanied by vast changes in the expression of genes appeared from the studies of 40 homoeologous genes in different organs of *G. hirsutum* (Adams, Cronn, Percifield and Wendel, 2003). Almost one-third of the studied genes demonstrated changes in expression towards a significant increase in the activity of one of the homoeologs and a decrease in the expression of the other. Special attention should be given to the genes that demonstrated organ-specific expression: while one of the genes in the homoeologous pair expressed itself in the organs of one type, the other gene was active only in the other organs (Adams, Cronn, Percifield and Wendel, 2003).

When studying the activity of the duplicated genes, it was also established that patterns of the homoeologous copies' expression were environment-sensitive. It was shown that the homoeologous genes of *G. hirsutum* demonstrate different levels of expression in different tissues under the influence of such abiotic stresses as an increase or decrease in temperature, deficiency or excess of water, and an increased content of salts (Liu and Adams, 2007; Dong and Adams, 2011). Probably, the differential expression of homoeologous genes in response to a stress or an environmental signal may be a factor that facilitates the preservation of the duplicated genes' functional state in the polyploid organism.

3.2. Expression of genes in synthetic cotton hybrids

The changes caused by distant hybridization at the early stages of allopolyploid organism formation should differ from those changes that took place in its subsequent evolution. In order to differentiate these changes, in a number of articles the expression of homoeologous genes was studied when comparing the synthetically created allopolyploids (or F1 hybrids) with the natural allopolyploid cotton species (Adams et al., 2003; Adams, Percifield and Wendel, 2004; Flagel, Udall, Nettleton and Wendel, 2008; Chaudhary et al., 2009). For example, a comparison of gene activity in the F1 hybrid, artificially produced by crossing *G. arboreum* and *G. raimondii,* and that in the natural *G. hirsutum* allopolyploid has shown that about 24 % of the genes with differential expression have demonstrated a similar change in expression in the F1 hybrid and in the natural allopolyploid, if compared with the parent forms (Flagel et al., 2008). The remaining 76 % of the genes of *G. hirsutum* with the expression changed in comparison with *G. arboreum* and *G. raimondii* could be determined by both the accumulated mutations and by the sub- or non-functionality of the duplicated genes (Flagel et al., 2008).

Thus, the merging of genomes plays an important, but only partial role in changing the pattern of expression of the homoeologous genes in the genus *Gossypium*, while the level of expression and tissue specificity of the genes demonstrate that specific patterns of the homoeologous genes expression can appear in both *de novo* created synthetic hybrids and remain in natural allopolyploids.

3.3. Dominance of allopolyploid cotton subgenomes

In allopolyploid plant forms resulting from interspecific hybridization, one of the parent subgenomes is dominant as a rule, i.e., it preserves the expression of homoeologous genes at a level similar to that of the genes' activity in the parent organism in relation to other homoeologs (Wang et al., 2006; Rapp, Udall and Wendel, 2009; Buggs et al.,

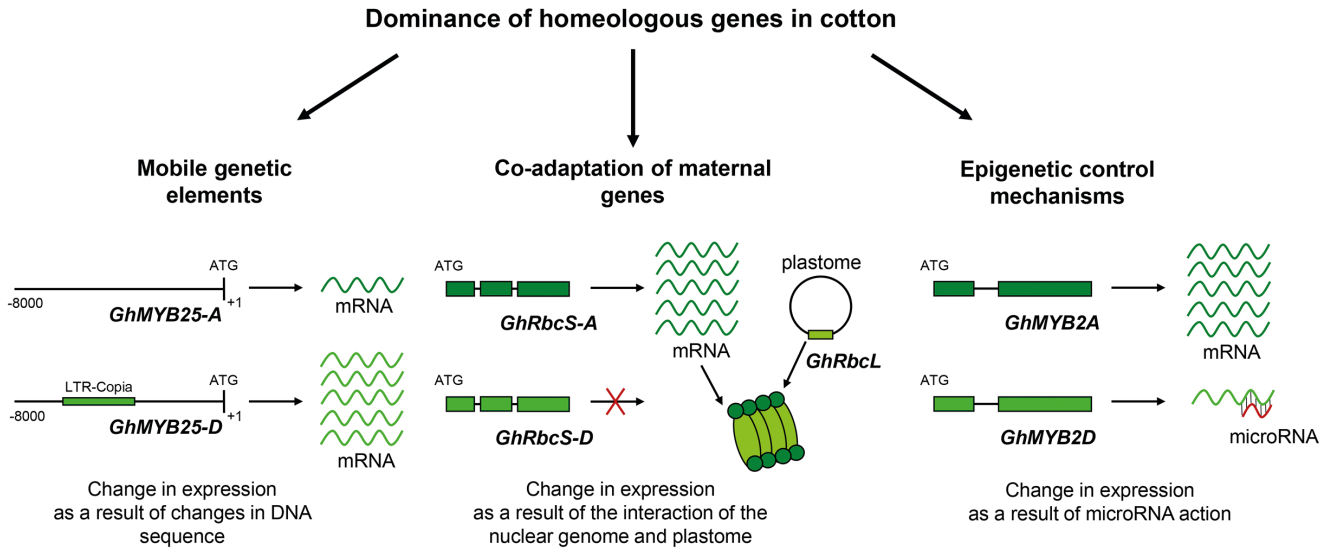

Fig. 3. Regulation of expression and dominance of homoeologous genes of allotetraploid cotton. Explanation is given in the text.

2010; Chang et al., 2010; Flagel and Wendel, 2010; Woodhouse et al., 2010; Schnable, Springer and Freeling, 2011; Tang et al., 2012). Interestingly, such an expression profile is characteristic of the genes with both increased and decreased expression levels, so that the same diploid parent genome can be either dominant or recessive, depending on the particular combination (Rapp et al., 2009).

The dominance of a subgenome can manifest itself in a situation when the maternal genes of the nuclear genome and the plastome co-adapted to each other can be expressed and the paternal homoeologous copies silenced. An example of this kind has been described for the *RuBisCO* genes (Gong et al., 2012). The *RuBisCO* enzyme (4.1.1.39; ribulose-1.5-bisphosphate carboxylase/oxygenase), which catalyzes the addition of CO_2 to ribulose-1.5-bisphosphate in the Calvin cycle, consists of small subunits (SSU) encoded by nuclear genes, and of large subunits (LSU) encoded by plastome genes (Fig. 3) (Rodermel et al., 1996). The genes of the maternal A genome were shown to dominate in polyploid cotton, i.e., both the nuclear genes of the small *rbcS* subunits, and the genes of the large *rbcL* subunits transmitted from the mother, which may demonstrate co-evolution of genes of large and small subunits (Gong et al., 2012).

The subgenome dominance could also be associated with natural selection, which in the course of evolution eliminates various problems with regulating trait manifestation, caused by the fusion of genomes (Yoo, Szadkowski and Wendel, 2013). For instance, the shift in the expression of genes involved in fiber formation in allotetraploid cotton is directed towards the D genome as a rule (Flagel et al., 2008; Hovav et al., 2008; Guan, Song and Chen, 2014). Since the D genome donors are incapable of producing fiber, it is quite likely that negative regulators, such as microRNAs and transcriptional repressors, suppress the expression of genes related to fiber formation in the D genome, compared to the A genome.

The homoeologous regulatory genes of the *R2R3-MYB* type — *GhMYB2A* and *GhMYB2D* — could serve as an example. They are homoeologs of the *A. thaliana GLABROUS1* (*GL1)* gene involved in trichome formation (Fig. 3) (Wang et al., 2004; Ishida, Kurata, Okada and Wada, 2008; Pesch and Hülskamp, 2009; Guan and Pang et al., 2014). It has been demonstrated that more mRNAs of the *GhMYB2D* gene than of *GhMYB2A* are synthesized during the initiation of cotton fiber formation (Guan and Pang et al., 2014). However, only *GhMYB2A* is involved in the process of fiber formation, since the products of *GhMYB2D* are the targets for miR828 and miR858 microRNAs (Fig. 3) (Pang et al., 2009; Guan and Pang et al., 2014).

Besides, it turned out that in *A. thaliana gl1*-mutants, i.e., mutants incapable of forming trichomes, the overexpression of *GhMYB2A* restores the mutant phenotype (Guan and Pang et al., 2014). Normally, the overexpression of *GhMYB2D* does not restore the *gl1* phenotype, but in the case of the miR828-binding site mutation, trichomes development is restored in *gl1*-mutants (Guan and Pang et al., 2014). Thus, these studies suppose not only functional divergence between *GhMYB2A* and *GhMYB2D* in cotton, but also an important role of microRNAs in the process of fiber formation.

In total, the studies show that gene expression in polyploid cotton changes in comparison with its diploid predecessors, and unequal expression of one of the two homoeologs is a rule rather than an exception.

Conclusion

The review summarizes the results of works devoted to the studies on phylogenetic relationships in the genus *Gossypium*, evolution of genomes within this genus, regulation of homoeologous genes' expression and dominance of allotetraploid cotton subgenomes. Despite the large amount of available data, the status of individual taxa in the *Gossypium* genus is still unclear, especially within the *Gossypium* and *Sturtia* subgenera, due to their poor representation in collections, inaccessibility of the territories where representatives of these taxa occur and the tendency to produce interspecific hybrids. Studies of the distribution patterns of repeating DNA elements, such as MGEs, can shed light on the evolution of genomes in the genus *Gossypium* and on the time of their divergence. On the other hand, the analysis of MGEs polymorphism could help to reveal the genes that control fiber development in cotton. Further studies, combined with the available data, will offer ample opportunities for producing cotton varieties with the desired properties.

Acknowledgments

The present review was carried out within VIR project No. 0481-2019-0001.

References

Adams, K. L., Cronn, R., Percifield, R., and Wendel, J. F. 2003. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proceedings of the National Academy of Sciences USA* 100(8):4649–4654. https://doi.org/10.1073/pnas.0630618100

Adams, K. L., Percifield, R., and Wendel, J. F. 2004. Organ-specific silencing of duplicated genes in a newly synthesized cotton allotetraploid. *Genetics* 168(4):2217–2226. https://doi.org/10.1534/genetics.104.033522

Alix, K., Gérard, P. R., Schwarzacher, T., and Heslop-Harrison, J. S. (Pat). 2017. Polyploidy and interspecific hybridization: Partners for adaptation, speciation and evolution in plants. *Annals of Botany* 120(2):183–194. https://doi.org/10.1093/aob/mcx079

Bennetzen, J. L., Ma, J., and Devos, K. M. 2005. Mechanisms of recent genome size variation in flowering plants. *Annals of Botany* 95(1):127–132. https://doi.org/10.1093/aob/mci008

Birchler, J. A., Riddle, N. C., Auger, D. L., and Veitia, R. A. 2005. Dosage balance in gene regulation: Biological implications. *Trends in Genetics* 21(4):219–226. https://doi.org/10.1016/j.tig.2005.02.010

Bowers, J. E., Chapman, B. A., Rong, J., and Paterson, A. H. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422(6930):433–438. https://doi.org/10.1038/nature01521

Brubaker, C. L., Paterson, A. H., and Wendel, J. F. 1999. Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. *Genome* 42(2):184–203. https://doi.org/10.1139/g98-118

Buggs, R. J. A., Chamala, S., Wu, W., Gao, L., May, G. D., Schnable, P. S., Soltis, D. E., Soltis, P. S., and Barbazuk, W. B. 2010. Characterization of duplicate gene evolution in the recent natural allopolyploid *Tragopogon miscellus* by next-generation sequencing and Sequenom iPLEX MassARRAY genotyping. *Molecular Ecology* 19:132–146. https://doi.org/10.1111/j.1365-294X.2009.04469.x

Campbell, B. T., Saha, S., Percy, R., Frelichowski, J., Jenkins, J. N., Park, W., Mayee, C. D., Gotmare, V., Dessauw, D., Giband, M., Du, X., Jia, Y., Constable, G., Dillon, S., Abdurakhmonov, I. Y., Abdukarimov, A., Rizaeva, S. M., Adullaev, A., Barroso, P. A. V., Pádua, J. G., Hoffmann, L. V., and Podolnaya, L. 2010. Status of the global cotton germplasm resources. *Crop Science* 50(4):1161–1179. https://doi.org/10.2135/cropsci2009.09.0551

Chang, P. L., Dilkes, B. P., McMahon, M., Comai, L., and Nuzhdin, S. V. 2010. Homoeolog-specific retention and use in allotetraploid *Arabidopsis suecica* depends on parent of origin and network partners. *Genome Biology* 11(12):R125. https://doi.org/10.1186/gb-2010-11-12-r125

Chaudhary, B., Flagel, L., Stupar, R. M., Udall, J. A., Verma, N., Springer, N. M., and Wendel, J. F. 2009. Reciprocal silencing, transcriptional bias and functional divergence of homeologs in polyploid cotton (*Gossypium*). *Genetics* 182(2):503–517. https://doi.org/10.1534/genetics.109.102608

Chen, Z. J. 2007. Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annual Review of Plant Biology* 58(1):377–406. https://doi.org/10.1146/annurev.arplant.58.032806.103835

Cronn, R. C., Zhao, X., Paterson, A. H., and Wendel, J. F. 1996. Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. *Journal of Molecular Evolution* 42(6):685–705. https://doi.org/10.1007/BF02338802

Cronn, R. and Wendel, J. F. 2003. Cryptic trysts, genomic mergers, and plant speciation. *New Phytologist* 161:133–142. https://doi.org/10.1046/j.1469-8137.2003.00947.x

Cui, X., Liu, F., Liu, Y., Zhou, Z., Wang, C., Zhao, Y., Wang, C., Zhao, Y., Meng, F., Wang, X., Cai, X., Wang, Y., Peng, R., and Wang, K. 2016. Screening and chromosome localization of two cotton BAC clones. *Comparative Cytogenetics* 10(1):1–15. https://doi.org/10.3897/CompCytogen.v10i1.5304

Davie, H. J. 1933. Cytological studies in the Malvaceae and certain related families. *Journal of Genetics* 28(1):33–67. https://doi.org/10.1007/BF02981768

Denham, H. J. 1924. The cytology of the cotton plant. I. Microspore formation in sea island cotton. *Annals of Botany* 38(151):407–432. https://doi.org/10.1093/oxfordjournals.aob.a089908

Dillehay, T. D., Rossen, J., Andres, T. C., and Williams, D. E. 2007. Preceramic adoption of peanut, squash, and cotton in Northern Peru. *Science* 316(5833):1890–1893. https://doi.org/10.1126/science.1141395

Dong, S. and Adams, K. L. 2011. Differential contributions to the transcriptome of duplicated genes in response to abiotic stresses in natural and synthetic polyploids. *New Phytologist* 190(4):1045–1057. https://doi.org/10.1111/j.1469-8137.2011.03650.x

Endrizzi, J. E., Turcotte, E. L., and Kohel, R. J. 1985. Genetics, cytology, and evolution of *Gossypium*. *Advances in Genetics* 23: 271–375. https://doi.org/10.1016/S0065-2660(08)60515-5

Fang, D. D. 2018. Cotton fiber: physics, chemistry and biology. Springer. https://doi.org/10.1007/978-3-030-00871-0

Flagel E. L. and Wendel F. J. 2010. Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation. *New Phy-*

tologist 186(1):184–193. https://doi.org/10.1111/j.1469-8137.2009.03107.x

Flagel, L., Udall, J., Nettleton, D., and Wendel, J. 2008. Duplicate gene expression in allopolyploid *Gossypium* reveals two temporally distinct phases of expression evolution. *BMC Biology* 6(1):16. https://doi.org/10.1186/1741-7007-6-16

Fryxell, P. A. 1968. A classification of *Gossypium* L. (Malvaceae). *Taxon* 18(5):585–591. https://doi.org/10.2307/1218405

Fryxell, P. A. 1971. Phenetic analysis and the phylogeny of the diploid species of *Gossypium* L. (Malvaceae). *Evolution* 25:554–562. https://doi.org/10.2307/2407355

Fryxell, P. A. 1978. The natural history of the cotton tribe (Malvaceae, tribe Gossypieae). *Texas A and M University Press.*

Fryxell, P. A. 1992. A revised taxonomic interpretation of *Gossypium* L. (Malvaceae). *Rheedea* 2:108–165.

Fryxell, P. A., Craven, L. A., and Stewart, J. M. 1992. A revision of *Gossypium* sect. *Grandicalyx* (Malvaceae), including the description of six new species. *Systematic Botany* 17(1):91–114. https://doi.org/10.2307/2419068

Gallagher, J. P., Grover, C. E., Rex, K., Moran, M., and Wendel, J. F. 2017. A new species of cotton from Wake Atoll, *Gossypium stephensii* (Malvaceae). *Systematic Botany* 42(1):115–123. https://doi.org/10.1600/036364417x694593

Geever, R. F. I., Katterman, F. R. H., and Endrizzi, J. E. 1989. DNA hybridization analyses of a *Gossypium* allotetraploid and two closely related diploid species. *Theoretical and Applied Genetics* 77:553–559. https://doi.org/10.1007/BF00274278

Golubets, V. 1991. Wild cottons of Africa and of the Arabian Peninsula. Plant industry, breeding and genetics of industrial plants. *Bulletin of applied botany, genetics and plant breeding* 144:26–30.

Gong, L., Salmon, A., Yoo, M. J., Grupp, K. K., Wang, Z., Paterson, A. H., and Wendel, J. F. 2012. The cytonuclear dimension of allopolyploid evolution: an example from cotton using rubisco. *Molecular Biology and Evolution* 29(10):3023–3036. https://doi.org/10.1093/molbev/mss110

Grover, C. E., Zhu, X., Grupp, K. K., Jareczek, J. J., Gallagher, J. P., Szadkowski, E., Seijo, J. G., and Wendel, J. F. 2014. Molecular confirmation of species status for the allopolyploid cotton species, *Gossypium ekmanianum* Wittmack. *Genetic Resources and Crop Evolution* 62(1):103–114. https://doi.org/10.1007/s10722-014-0138-x

Grover, C. E., Kim, H. R., Wing, R. A., Paterson, A. H., and Wendel, J. F. 2004. Incongruent patterns of local and global genome size evolution in cotton. *Genome Research* 14(8):1474–1482. https://doi.org/10.1101/gr.2673204

Grover, C. E, Arick II, M. A., Thrash, A., Conover, J. L., Sanders, W. S., Peterson, D. G., Frelichowski, J. E., Scheffler, J. A., Scheffler, B. E., and Wendel, J. F. 2018. Insights into the evolution of the New World diploid cottons (*Gossypium*, subgenus *Houzingenia*) based on genome sequencing. *Genome Biology and Evolution* 11(1):53–71. https://doi.org/10.1093/gbe/evy256

Guan, X., Pang, M., Nah, G., Shi, X., Ye, W., Stelly, D. M., and Chen, Z. J. 2014. MiR828 and miR858 regulate homoeologous MYB2 gene functions in *Arabidopsis* trichome and cotton fibre development. *Nature Communications* 5:3050. https://doi.org/10.1038/ncomms4050

Guan, X., Song, Q., and Chen, Z. J. 2014. Polyploidy and small RNA regulation of cotton fiber development. *Trends in Plant Science* 19(8):516–528. https://doi.org/10.1016/j.tplants.2014.04.007

Hawkins, J. S., Kim, H. R., Nason, J. D., Wing, R. A., and Wendel, J. F. 2006. Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. *Genome Research* 16(10):1252–1261. https://doi.org/10.1101/gr.5282906

Hendrix, B. and Stewart, J. M. 2005. Estimation of the nuclear DNA content of *Gossypium* species. *Annals of Botany* 95(5):789–797. https://doi.org/10.1093/aob/mci078

Hovav, R., Udall, J. A., Chaudhary, B., Rapp, R., Flagel, L., and Wendel, J. F. 2008. Partitioned expression of duplicated genes during development and evolution of a single cell in a polyploid plant. *Proceedings of the National Academy of Sciences USA* 105(16):6191–6195. https://doi.org/10.1073/pnas.0711569105

Ishida, T., Kurata, T., Okada, K., and Wada, T. 2008. A genetic regulatory network in the development of trichomes and root hairs. *Annual Review of Plant Biology* 59(1):365–386. https://doi.org/10.1146/annurev.arplant.59.032607.092949

Jiang, C.-X., Wright, R. J., El-Zik, K. M., and Paterson, A. H. 1998. Polyploid formation created unique avenues for response to selection in *Gossypium* (cotton). *Proceedings of the National Academy of Sciences USA* 95:4419–4424. https://doi.org/10.1073/pnas.95.8.4419

Jiao, Y. and Paterson, A. H. 2014. Polyploidy-associated genome modifications during land plant evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369(1648):20130355. https://doi.org/10.1098/rstb.2013.0355

Kim, H. J. 2015. Fiber Biology. In: Cotton, 2nd ed., Agron. Monogr. 57. ASA, CSSA, and SSSA, Madison, pp. 97–128. https://doi.org/10.2134/agronmonogr57.2013.0022

Lawrence, W. J. C. 1931. The secondary association of chromosomes. *Cytologia* 2(4):352–384. https://doi.org/10.1508/cytologia.2.352

Lemeshev, N. 1991. The problems of origin of allotetraploid cotton species. Plant industry, breeding and genetics of industrial plants. *Bulletin of applied botany, genetics and plant breeding* 144:3–9.

Li, F., Fan, G., Lu, C., Xiao, G., Zou, C., Kohel, R. J., Ma, Z., Shang, H., Ma, X., Wu, J., Liang, X., Huang, G., Percy, R. G., Liu, K., Yang, W., Chen, W., Du, X., Shi, C., Yuan, Y., Ye, W., Liu, X., Zhang, X., Liu, W., Wei, H., Wei, S., Huang, G., Zhang, X., Zhu, S., Zhang, H., Sun, F., Wang, X., Liang, J., Wang, J., He, Q., Huang, L., Wang, J., Cui, J., Song, G., Wang, K., Xu, X., Yu, J. Z., Zhu, Y., and Yu, S. 2015. Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nature Biotechnology* 33(5):524–530. https://doi.org/10.1038/nbt.3208

Li, F., Fan, G., Wang, K., Sun, F., Yuan, Y., Song, G., Li, Q., Ma, Z., Lu, C., Zou, C., Chen, W., Liang, X., Shang, H., Liu, W., Shi, C., Xiao, G., Gou, C., Ye, W., Xu, X., Zhang, X., Wei, H., Li, Z., Zhang, G., Wang, J., Liu, K., Kohel, R. J., Percy, R. G., Yu, J. Z., Zhu Y. X., Wang J., and Yu, S. 2014. Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nature Genetics* 46(6):567–572. https://doi.org/10.1038/ng.2987

Liu, X., Zhao, B., Zheng, H. J., Hu, Y., Lu, G., Yang, C. Q., Chen, J. D., Chen, J. J., Chen D. Y., Zhang, L., Zhou, Y., Wang, L. J., Guo, W. Z., Bai, Y. L., Ruan, J. X., Shangguan, X. X., Mao, Y. B., Shan, C. M., Jiang, J. P., Zhu, Y. Q., Jin, L., Kang, H., Chen, S. T., He, X. L., Wang, R., Wang, Y. Z., Chen, J., Wang, L. J., Yu, S. T., Wang, B. Y., Wei, J., Song, S. C., Lu, X. Y., Gao, Z. C., Gu, W. Y., Deng, X., Ma, D., Wang, S., Liang, W. H., Fang, L., Cai, C. P., Zhu, X. F., Zhou, B. L., Chen, Z. J., Xu, S. H., Zhang, Y. G., Wang, S. Y., Zhang, T. Z., Zhao, G. P., and Chen, X. Y. 2015. *Gossypium barbadense* genome sequence provides insight into the evolution of extra-long staple fiber and specialized metabolites. *Scientific Reports* 5:1–14. https://doi.org/10.1038/srep14139

Liu, Z. and Adams, K. L. 2007. Expression partitioning between genes duplicated by polyploidy under abiotic stress and during organ development. *Current Biology* 17(19):1669–1674. https://doi.org/10.1016/j.cub.2007.08.030

Lu, H., Cui, X., Liu, Z., Liu, Y., Wang, X., Zhou, Z., Cai, X., Zhang, Z., Guo, X., Hua, J., Ma, Z., Wang, X., Zhang, J., Zhang, H., Liu, F., and Wang, K. 2018. Discovery and annotation of a novel transposable element family in *Gossypium*. *BMC Plant Biology* 18(1):307. https://doi.org/10.1186/s12870-018-1519-7

Osborn, T. C., Pires, J. C., Birchler, J. A., Auger, D. L., Chen, Z. J., Lee, H.-S., Comai, L., Madlung, A., Doerge, R. W., Colot, V., and Martienssen, R. A. 2003. Understanding mechanisms of novel gene expression in polyploids. *Trends in Genetics* 19(3):141–147. https://doi.org/10.1016/S0168-9525(03)00015-5

Ozkan, H., Tuna, M., Kilian, B., Mori, N., and Ohta, S. 2010. Genome size variation in diploid and tetraploid wild wheats. *AoB Plants* plq015. https://doi.org/10.1093/aobpla/plq015

Page, J. T., Huynh, M. D., Liechty, Z. S., Grupp, K., Stelly, D., Hulse, A. M., Ashrafi, H., Deynze, A. V., Wendel, J. F., and Udall, J. A. 2013. Insights into the evolution of cotton diploids and polyploids from whole-genome re-sequencing. *G3: GENES, GENOMES, GENETICS* 3(10):1809–1818. https://doi.org/10.1534/g3.113.007229

Panchy, N., Lehti-Shiu, M. D., and Shiu, S.-H. 2016. Evolution of gene duplication in plants. *Plant Physiology* 171(4):2294–2316. https://doi.org/10.1104/pp.16.00523

Pang, M., Woodward, A. W., Agarwal, V., Guan, X., Ha, M., Ramachandran, V., Chen, X., Triplett, B. A., Stelly, D. M., and Chen, Z. J. 2009. Genome-wide analysis reveals rapid and dynamic changes in miRNA and siRNA sequence and expression during ovule and fiber development in allotetraploid cotton (*Gossypium hirsutum* L.). *Genome Biology* 10(11):R122. https://doi.org/10.1186/gb-2009-10-11-r122

Paterson, A. H. 2009. Genetics and genomics of cotton. Springer. https://doi.org/10.1007/978-0-387-70810-2

Paterson, A. H., Wendel, J. F., Gundlach, H., Guo, H., Jenkins, J., Jin, D., Llewellyn, D., Showmaker, K. C., Shu, S., Udall, J., Yoo, M. J., Byers, R., Chen, W., Doron-Faigenboim, A., Duke, M. V., Gong, L., Grimwood, J., Grover, C., Grupp, K., Hu, G., Lee, T. H., Li, J., Lin, L., Liu, T., Marler, B. S., Page, J. T., Roberts, A. W., Romanel, E., Sanders, W. S., Szadkowski, E., Tan, X., Tang, H., Xu, C., Wang, J., Wang, Z., Zhang, D., Zhang, L., Ashrafi, H., Bedon, F., Bowers, J. E., Brubaker, C. L., Chee, P. W., Das, S., Gingle, A. R., Haigler, C. H., Harker, D., Hoffmann, L. V., Hovav, R., Jones, D. C., Lemke, C., Mansoor, S., Rahman, M. U., Rainville, L. N., Rambani, A., Reddy, U. K., Rong, J. K., Saranga, Y., Scheffler, B. E., Scheffler, J. A., Stelly, D. M., Triplett, B. A., Deynze, A. V., Vaslin, M. F. S., Waghmare, V. N., Walford, S. A., Wright, R. J., Zaki, E. A., Zhang, T., Dennis, E. S., Mayer, K. F. X., Peterson, D. G., Rokhsar, D. S., Wang, X., and Schmutz, J. 2012. Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature* 492(7429):423–427. https://doi.org/10.1038/nature11798

Pesch, M. and Hülskamp, M. 2009. One, two, three...models for trichome patterning in *Arabidopsis*? *Current Opinion in Plant Biology* 12(5):587–592. https://doi.org/10.1016/j.pbi.2009.07.015

Phuphathanaphong, L. 2006. *Thepparatia* (Malvaceae), a new genus from Thailand. *Thai Forest Bulletin (Botany)* 34:195–200.

Podolnaya, L. P. 1991. Seedlings morphology of wild species of the genus *Gossypium*. Plant industry, breeding and genetics of industrial plants. *Bulletin of applied botany, genetics and plant breeding* 144:18–25.

Qin, Y.-M., Hu, C.-Y., Pang, Y., Kastaniotis, A. J., Hiltunen, J. K., and Zhu, Y.-X. 2007. Saturated very-long-chain fatty acids promote cotton fiber and *Arabidopsis* cell elongation by activating ethylene biosynthesis. *The Plant Cell Online* 19(11):3692–3704. https://doi.org/10.1105/tpc.107.054437

Rapp, R. A., Udall, J. A., and Wendel, J. F. 2009. Genomic expression dominance in allopolyploids. *BMC Biology* 7(1):18. https://doi.org/10.1186/1741-7007-7-18

Renny-Byfield, S., Gallagher, J. P., Grover, C. E., Szadkowski, E., Page, J. T., Udall, J. A., Wang, X., Paterson, A. H., and Wendel, J. F. 2014. Ancient gene duplicates in *Gossypium* (cotton) exhibit near-complete expression divergence. *Genome Biology and Evolution* 6(3):559–571. https://doi.org/10.1093/gbe/evu037

Renny-Byfield, S., Gong, L., Gallagher, J. P., and Wendel, J. F. 2015. Persistence of subgenomes in paleopolyploid cotton after 60 my of evolution. *Molecular Biology and Evolution* 32(4):1063–1071. https://doi.org/10.1093/molbev/msv001

Riddle, N. C. and Birchler, J. A. 2003. Effects of reunited diverged regulatory hierarchies in allopolyploids and species hybrids. *Trends in Genetics* 19(11):593–597. https://doi.org/10.1016/j.tig.2003.09.009

Rodermel, S., Haley, J., Jiang, C. Z., Tsai, C. H., and Bogorad, L. 1996. A mechanism for intergenomic integration: abundance of ribulose bisphosphate carboxylase small-subunit protein influences the translation of the large-subunit mRNA. *Proceedings of the National Academy of Sciences USA* 93(9):3881–3885. https://doi.org/10.1073/pnas.93.9.3881

Rong, J., Feltus, F. A., Liu, L., Lin, L., and Paterson, A. H. 2010. Gene copy number evolution during tetraploid cotton radiation. *Heredity* 105(5):463–472. https://doi.org/10.1038/hdy.2009.192

SanMiguel, P. and Bennetzen, J. L. 1998. Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. *Annals of Botany* 82:37–44. https://doi.org/10.1006/anbo.1998.0746

Sattler, M. C., Carvalho, C. R., and Clarindo, W. R. 2016. The polyploidy and its key role in plant breeding. *Planta* 243(2):281–296. https://doi.org/10.1007/s00425-015-2450-x

Schnable, J. C., Springer, N. M., and Freeling, M. 2011. Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proceedings of the National Academy of Sciences USA* 108(10):4069–4074. https://doi.org/10.1073/pnas.1101368108

Seelanan, T., Brubaker, C. L., Stewart, J. M., Craven, L. A., Wendel, J. F., and Liston, A. 1999. Molecular systematics of Australian *Gossypium* section *Grandicalyx* (Malvaceae). *Systematic Botany* 24(2):183–208. https://doi.org/10.2307/2419548

Seelanan, T., Schnabel, A., and Wendel, J. F. 1997. Congruence and consensus in the cotton tribe (Malvaceae). *Systematic Botany* 22(2):259–290. https://doi.org/10.2307/2419457

Senchina, D. S., Alvarez, I., Cronn, R. C., Liu, B., Rong, J., Noyes, R. D., Paterson, A. H., Wing, R. A., Wilkins, T. A., and Wendel, J. F. 2003. Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. *Molecular Biology and Evolution* 20(4):633–643. https://doi.org/10.1093/molbev/msg065

Shi, Y.-H., Zhu, S.-W., Mao, X.-Z., Feng, J.-X., Qin, Y.-M., Zhang, L., Cheng, J., Wei, L. P., Wang, Z.-Y., and Zhu, Y.-X. 2006. Transcriptome profiling, molecular biological, and

physiological studies reveal a major role for ethylene in cotton fiber cell elongation. *The Plant Cell* 18(3):651–664. https://doi.org/10.1105/tpc.105.040303

Skovsted, A. 1933. Cytological studies in cotton. I. The mitosis and the meiosis in diploid and triploid Asiatic cotton. *Annals of Botany* 47(186):227–251. https://doi.org/10.1093/oxfordjournals.aob.a090380

Skovsted, A. 1934. Cytological studies in cotton. II. Two interspecific hybrids between Asiatic and new world cottons. *Journal of Genetics* 28(3): 407–424. https://doi.org/10.1007/BF02981763

Small, R. L. and Wendel, J. F. 1999. The mitochondrial genome of allotetraploid cotton (*Gossypium* L.). *The Journal of Heredity* 90(1):251–253. https://doi.org/10.1093/jhered/90.1.251

Taliercio, E. W. and Boykin, D. 2007. Analysis of gene expression in cotton fiber initials. *BMC Plant Biology* 7(1):22. https://doi.org/10.1186/1471-2229-7-22

Tang, H., Woodhouse, M. R., Cheng, F., Schnable, J. C., Pedersen, B. S., Conant, G., Wang, X., Freeling, M., and Pires, J. C. 2012. Altered patterns of fractionation and exon deletions in *Brassica rapa* support a two-step model of paleohexaploidy. *Genetics* 190(4):1563–1574. https://doi.org/10.1534/genetics.111.137349

Tiwari, R. S., Zhang, J., and Stewart, J. Mac. 2014. Molecular diversity based on AFLP markers and possible natural hybridization among the Australian arid zone *Gossypium* species. *Australian Journal of Crop Science* 8(5):674–679.

Ulloa, M. 2014. The diploid D genome cottons (*Gossypium* spp.) of the New World. IntechOpen. https://doi.org/10.5772/58387

Vollesen, K. 1987. The native species of *Gossypium* (Malvaceae) in Africa, Arabia and Pakistan. *Kew Bulletin* 42(2):337–349. https://doi.org/10.2307/4109688

Walford, S. A., Wu, Y., Llewellyn, D. J., and Dennis, E. S. 2011. GhMYB25-like: A key factor in early cotton fibre development. *Plant Journal* 65(5):785–797. https://doi.org/10.1111/j.1365-313X.2010.04464.x

Wang, J., Tian, L., Lee, H. S., Wei, N. E., Jiang, H., Watson, B., Madlung, A., Osborn, T. C., Doerge, R. W., Comai, L., and Chen, Z. J. 2006. Genomewide nonadditive gene regulation in arabidopsis allotetraploids. *Genetics* 172(1):507–517. https://doi.org/10.1534/genetics.105.047894

Wang, K., Huang, G., and Zhu, Y. 2016. Transposable elements play an important role during cotton genome evolution and fiber cell development. *Science China Life Sciences* 59(2):112–121. https://doi.org/10.1007/s11427-015-4928-y

Wang, K., Wang, Z., Li, F., Ye, W., Wang, J., Song, G., Yue, Z., Cong, L., Shang, H., Zhu, S., Zou, C., Li, Q., Yuan, Y., Lu, C., Wei, H., Gou, C., Zheng, Z., Yin, Y., Zhang, X., Liu, K., Wang, B., Song, C., Shi, N., Kohel, R. J., Percy, R. G., Yu, J. Z., Zhu, Y. X., Cang, J., and Yu, S. 2012. The draft genome of a diploid cotton *Gossypium raimondii*. *Nature Genetics* 44(10):1098–1103. https://doi.org/10.1038/ng.2371

Wang, Q. Q., Liu, F., Chen, X. S., Ma, X. J., Zeng, H. Q., and Yang, Z. M. 2010. Transcriptome profiling of early developing cotton fiber by deep-sequencing reveals significantly differential expression of genes in a fuzzless/lintless mutant. *Genomics* 96(6):369–376. https://doi.org/10.1016/j.ygeno.2010.08.009

Wang, S., Wang, J.-W., Yu, N., Li, C.-H., Luo, B., Gou, J.-Y., Wang, L.-J., and Chen, X.-Y. 2004. Control of plant trichome development by a cotton fiber MYB gene. *The Plant Cell* 16(9):2323–2334. https://doi.org/10.1105/tpc.104.024844

Webber, J. M. 1938. Cytology of twin cotton plants. *Journal of Agricultural Research* 57(2):155–160.

Wendel, J. F. 1989. New World tetraploid cottons contain Old World cytoplasm. *Proceedings of the National Academy of Sciences USA* 86:4132–4136. https://doi.org/10.1073/pnas.86.11.4132

Wendel, J. F. and Percival, A. E. 1990. Molecular divergence in the Galapagos Islands — Baja California species pair, *Gossypium klotzschianum* and *G . davidsonii* (Malvaceae). *Plant Systematics and Evolution* 171:99–115. https://doi.org/10.1007/BF00940598

Wendel, J. F. and Albert, V. A. 1992. Phylogenetics of the cotton genus (*Gossypium*): character-state weighted parsimony analysis of chloroplast-DNA restriction site data and its systematic and biogeographic implications. *Systematic Botany* 17(1):115–143. https://doi.org/10.2307/2419069

Wendel, J. F. and Grover, C. E. 2015. Taxonomy and evolution of the cotton genus, *Gossypium*. In: Cotton, 2nd ed., Agron. Monogr. 57. ASA, CSSA, and SSSA, Madison, pp. 25–44. https://doi.org/10.2134/agronmonogr57.2013.0020

Wendel, J. F., Olson, P. D., and Stewart, J. M. 1989. Genetic diversity, introgression, and independent domestication of Old World cultivated cottons. *American Journal of Botany* 76(12):1795. https://doi.org/10.2307/2444478

Wendel, J. F., Stewart, J. M., and Rettig, J. H. 1991. Molecular evidence for homoploid reticulate evolution among Australian species of *Gossypium*. *Evolution* 45(3):694. https://doi.org/10.2307/2409921

Wendel, J. F., Brubaker, C., Alvarez, I., Cronn, R., and Stewart, J. M. 2009. Genetics and genomics of cotton. Springer. https://doi.org/10.1007/978-0-387-70810-2

Wendel, J. F. and Cronn, R. C. 2003. Polyploidy and the evolutionary history of cotton. *Adavnce in Agronomy* 78:139–186. https://doi.org/10.1016/S0065-2113(02)78004-8

Woodhouse, M. R., Schnable, J. C., Pedersen, B. S., Lyons, E., Lisch, D., Subramaniam, S., and Freeling, M. 2010. Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homeologs. *PLoS Biology* 8(6):e1000409. https://doi.org/10.1371/journal.pbio.1000409

Wu, Y., Llewellyn, D. J., White, R., Ruggiero, K., Al-Ghazi, Y., and Dennis, E. S. 2007. Laser capture microdissection and cDNA microarrays used to generate gene expression profiles of the rapidly expanding fibre initial cells on the surface of cotton ovules. *Planta* 226(6):1475–1490. https://doi.org/10.1007/s00425-007-0580-5

Wu, Y., Machado, A. C., White, R. G., Llewellyn, D. J., and Dennis, E. S. 2006. Expression profiling identifies genes expressed early during lint fibre initiation in cotton. *Plant and Cell Physiology* 47(1):107–127. https://doi.org/10.1093/pcp/pci228

Wu, Z., Yang, Y., Huang, G., Lin, J., Xia, Y., and Zhu, Y. 2017. Cotton functional genomics reveals global insight into genome evolution and fiber development. *Journal of Genetics and Genomics* 44(11):511–518. https://doi.org/10.1016/j.jgg.2017.09.009

Xu, Z., Yu, J., Kohel, R. J., Percy, R. G., Beavis, W. D., Main, D., and Yu, J. Z. 2015. Distribution and evolution of cotton fiber development genes in the fibreless *Gossypium raimondii* genome. *Genomics* 106(1):61–69. https://doi.org/10.1016/j.ygeno.2015.03.002

Yoo, M. J., Szadkowski, E., and Wendel, J. F. 2013. Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity* 110(2):171–180. https://doi.org/10.1038/hdy.2012.94

Yu, J., Jung, S., Cheng, C. H., Ficklin, S. P., Lee, T., Zheng, P., Jones, D., Percy, R. G., and Main, D. 2014. CottonGen: A genomics, genetics and breeding database for cotton research. *Nucleic Acids Research* 42(D1):D1229–D1236. https://doi.org/10.1093/nar/gkt1064

Zhang, F., Zuo, K., Zhang, J., Liu, X., Zhang, L., Sun, X., and Tang, K. 2010. An L1 box binding protein, GbML1, interacts with GbMYB25 to control cotton fibre development. *Journal of Experimental Botany* 61(13):3599–3613. https://doi.org/10.1093/jxb/erq173

Zhang, T., Hu, Y., Jiang, W., Fang, L., Guan, X., Chen, J., Zhang, J., Saski, C. A., Scheffler, B. E., Stelly, D. M., Hulse-Kemp, A. M., Wan, Q., Liu, B., Liu, C., Wang, S., Pan, M., Wang, Y., Wang, D., Ye, W., Chang, L., Zhang, W., Song, Q., Kirkbride, R. C., Chen, X., Dennis, E., Llewellyn, D. J., Peterson, D. G., Thaxton, P., Jones, D. C., Wang, Q., Xu, X., Zhang, H., Wu, H., Zhou, L., Mei, G., Chen, S., Tian, Y., Xiang, D., Li, X., Ding, J., Zuo, Q., Tao, L., Liu, Y., Li, J., Lin, Y., Hui, Y., Cao, Z., Cai, C., Zhu X., Jiang, Z., Zhou, B., Guo, W., Li, R., and Chen, Z. J. 2015. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nature Biotechnology* 33(5):531–537. https://doi.org/10.1038/nbt.3207

Zhao, X. P., Si, Y., Hanson, R. E., Crane, C. F., Price, H. J., Stelly, D. M., Wendel, J. F., and Paterson, A. H. 1998. Dispersed repetitive DNA has spread to new genomes since polyploid formation in cotton. *Genome Research* 8(5):479–492. https://doi.org/10.1101/gr.8.5.479