SCIENCE CHINA
Life Sciences

THEMATIC ISSUE: Cotton genome plus
• REVIEW •

February 2016 Vol.59 No.2: 112–121
doi: 10.1007/s11427-015-4928-y

Transposable elements play an important role during cotton genome evolution and fiber cell development

Kun Wang[1], Gai Huang[3] & Yuxian Zhu[1,2*]

[1]*College of Life Sciences, Wuhan University, Wuhan 430072, China;*
[2]*Institute for Advanced Studies, Wuhan University, Wuhan 430072, China;*
[3]*State Key Laboratory of Protein and Plant Gene Research, College of Life Sciences, Peking University, Beijing 100871, China*

Received May 14, 2015; accepted July 20, 2015; published online December 17, 2015

Transposable elements (TEs) usually occupy largest fractions of plant genome and are also the most variable part of the structure. Although traditionally it is hallmarked as "junk and selfish DNA", today more and more evidence points out TE's participation in gene regulations including gene mutation, duplication, movement and novel gene creation via genetic and epigenetic mechanisms. The recently sequenced genomes of diploid cottons *Gossypium arboreum* (AA) and *Gossypium raimondii* (DD) together with their allotetraploid progeny *Gossypium hirsutum* (AtAtDtDt) provides a unique opportunity to compare genome variations in the *Gossypium* genus and to analyze the functions of TEs during its evolution. TEs accounted for 57%, 68.5% and 67.2%, respectively in DD, AA and AtAtDtDt genomes. The 1,694 Mb A-genome was found to harbor more LTR(long terminal repeat)-type retrotransposons that made cardinal contributions to the twofold increase in its genome size after evolution from the 775.2 Mb D-genome. Although the 2,173 Mb AtAtDtDt genome showed similar TE content to the A-genome, the total numbers of *LTR-gypsy* and *LTR-copia* type TEs varied significantly between these two genomes. Considering their roles on rewiring gene regulatory networks, we believe that TEs may somehow be involved in cotton fiber cell development. Indeed, the insertion or deletion of different TEs in the upstream region of two important transcription factor genes in At or Dt subgenomes resulted in qualitative differences in target gene expression. We suggest that our findings may open a window for improving cotton agronomic traits by editing TE activities.

transposable elements, *Gossypium* genus, genome evolution, fiber development

Citation: Wang, K., Huang, G., and Zhu, Y. (2016). Transposable elements play an important role during cotton genome evolution and fiber cell development. Sci China Life Sci 59, 112−121. doi: 10.1007/s11427-015-4928-y

INTRODUCTION

Early in 1950s, geneticist Barbara McClintock first noted the transposable elements (TEs) in maize (McClintock, 1948). She named them "controlling elements" because of her observation that it could alter the transcription of the gene nearby (McClintock, 1956). In the following decades, TEs did not get enough attention and even have been described as "junk" or "parasitic" DNA from most other scientists. However, recently mounting evidence was collected to suggest that TEs contributed to the gene regulatory network significantly (Feschotte, 2008). McClintock's discovery that TEs can control and modulate gene expression has now been supported by experiments across ever wider range of organisms.

TEs were categorized as class I (retrotransposons) and class II (DNA transposon) (Wicker et al., 2007). The retrotransposons, mainly including autonomous elements of LTR (long terminal repeat), DIRS_like (dictyostelium intermediate repeat sequence), PLEs (penelop-like elements), LINEs (long interspersed nuclear elements) and SINEs

*Corresponding author (email: zhuyx@whu.edu.cn)

(short interspersed nuclear elements) as well as non-autonomous elements, transpose by a "copy and paste" manner that used an RNA intermediate to make DNA copy. The LTR-type retrotransposons are rich in plants and divided into two major superfamilies: *gypsy* and *copia*, which differ in the order of the open reading frames (ORFs) that encode transposition related proteins. DNA transposons transpose a DNA intermediate within the genome that can also be classified in two subclasses (Wicker et al., 2007) One subclass "cut and paste" a double strand DNA in the genome, while the other subclass mobilize a single strand DNA with a "rolling circle" mechanism during DNA transposition (Li et al., 2009).

TE EFFECTS ON GENOME EVOLUTION

TE transposition together with hybridization, polyploidy or whole genome duplication (WGD), recombination, and horizontal gene transfer act in a concert to promote genome evolution (Oliver et al., 2013). It affects genome composition in two aspects. (i) Increase genome size: TEs often make up major parts of angiosperm plant genomes. TE-rich genomes such as maize (with up to 84% TE content (Schnable et al., 2009)) have been described as "gene islands surrounded by TE sea" (Oliver et al., 2013; Bennetzen et al., 2014; Tenaillon et al., 2010) and TE expansion plays important roles during genome expansion (Bennetzen et al., 2014; Piegu et al., 2006; Hawkins et al., 2006). For example, three LTR-type retrotransposon families in a wild rice became active within the last three million years that is responsible mainly for doubling its genome size (Piegu et al., 2006). (ii) Remove and rearrange genome fragment: unequal homologous recombination (HR) and illegitimate recombination (IR) are two major mechanisms of DNA removal (Bennetzen et al., 2007). TE also exerts roles in genome DNA removal. HR processes occurred between TEs of the same family were considered as one route for DNA translocation. Transposition of class II type TE with "cut and paste" mechanism may cause chromosome breaks that is utilized by IR during DNA removal as well.

TE EFFECTS ON GENE EXPRESSION

In addition to those macroscopic effects on the overall structure of genome, TEs cause widespread changes in the levels of gene expression. In recent years, plant biologists revealed numerous examples that TE mediated genotypic changes lead to different plant phenotype variations, which covers almost the whole plant life cycle, from seed formation to blossom and fruiting. Figure 1 shows several examples of TE-associated variations in higher plants. In maize, an insertion of LTR-type retrotransposon in *b1* exon resulted in its ectopic expression and anthocyanin accumulation kernels while a different TE insertion results in reduced and variegated expression of the gene with colors accumulated only in very few of the kernels (Figure 1A) (Selinger et al., 2001). In grape, the insertion of *Hatvine1-rrm*, a DNA TE in the *VvTFL1A* promoter caused up-regulation of the corresponding allele in reproductive and vegetative organs of the shoot apex of grape (Figure 1B) (Fernandez et al., 2010). In oranges, the insertion of LTR-type retrotransposon in the upstream of *Ruby* gene resulted in ecotopic and temperature-dependent expression of the gene in the flesh to produce so called "blood oranges" (Figure 1C) (Butelli et al., 2012). While the insertion of a *Gret1*, an LTR-type retrotransposon reduced *Vvmby1A* expression that leads to purple color loss in the grape mutant (Figure 1D) (Kobayashi et al., 2004). In tomato, an unusual 24.7 kb gene duplication event mediated by the LTR-type retrotransposon "*Rider*", which resulted in the increased *IQD12* (SUN locus) expression relative to that of the ancestral copy, culminating in an elongated fruit shape (Figure 1E) (Xiao et al., 2008). "Hopscotch", a transposable element inserted in a regulatory region of the maize domestication gene, *teosinte branched1* (*tb1*), acts as an enhancer of gene expression to increase apical dominance in maize compared to its progenitor, teosinte. Molecular evidence indicates that the "Hopscotch" insertion predates branch domestication of maize by at least 10,000 years (Figure 1F) (Studer et al., 2008). A CACTA-like transposable element (TE) represses the expression of *ZmCCT*, another maize domestication gene, to reduce photoperiod sensitivity, thus accelerating maize spread to long-day environments (Yang et al., 2013). In rice, an LTR-type retrotransposon insertion refunctionalize a "sleeping" R gene *pit* (disease resistance gene) in the rice genome to enhance fungal resistance in rice (Figure 1G) (Hayashi et al., 2009).

TE mediated changes in plants (and indeed in any eukaryote), that span from subtle quantitative effects on target gene expression to rewiring of regulatory networks and to new gene evolution, are categorized by either genetic or epigenetic functions. Genetic changes caused by TE insertion involves target gene mutation, structural modification, movement, evolution of novel gene and modulation of expression (Lisch, 2013). Figure 2 summarized some examples of TE-mediated epigenetic effects. In Figure 2A, the chromatin remodeling factor, deficient DNA methylation 1 (DDM1), is essential for DNA methylation, and its mutation has been reported to cause a profound loss of DNA methylation in *Arabidopsis*. For example, the DNA methylation of a SINE element silences its downstream gene locus of *FLOWERING WAGENINGEN (FWA)* in vegetative tissues. In *ddm1* mutant, the loss of DNA methylation in the SINE caused ectopic expression of *FWA* and results in a late flowering phenotype (Kinoshita et al., 2007). A second example was from several generations in *ddm1* mutant background. The loss of DNA methylation of a LINE activates TE transcription and produces antisense strand to its downstream gene *BONSAI*, finally resulting in the spread of DNA methylation and transcriptional silencing of this region, and severe dwarfing (Saze et al., 2007). In figure 2B, the

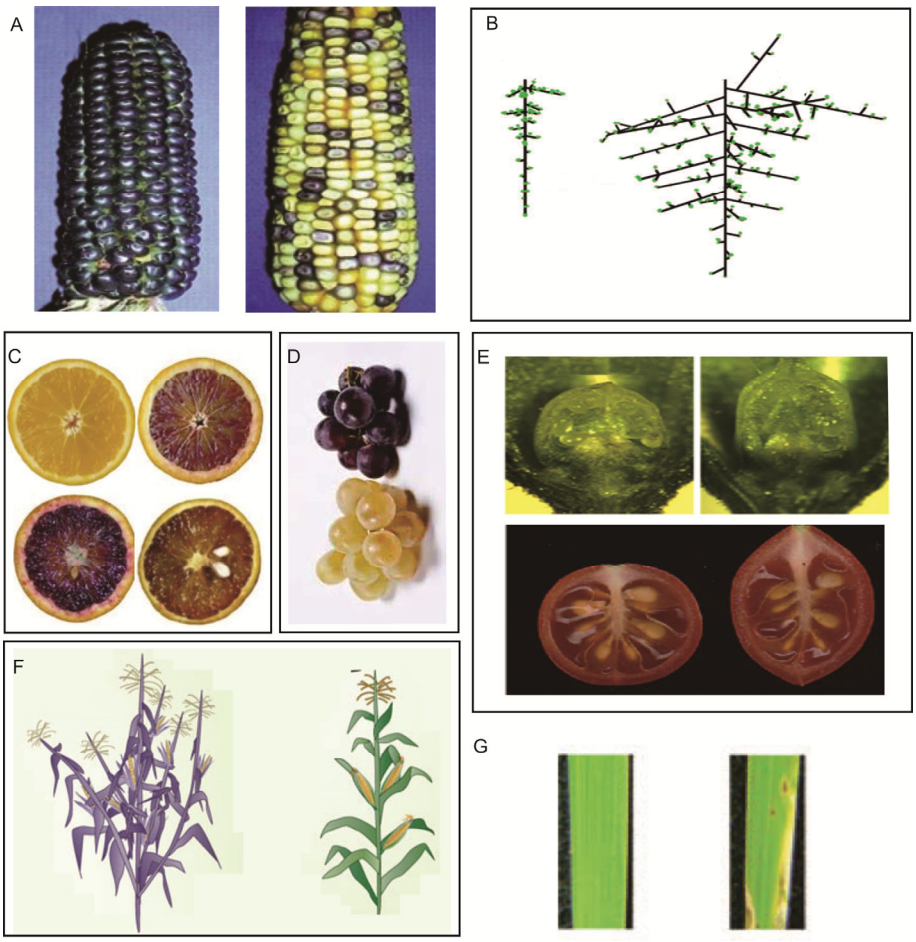

Figure 1 TEs mediated phenotypic changes in different plants. Insertion of LTR-type retrotransposon in *b1* exon results in an ectopic expression of the gene that accumulates anthocyanin in maize kernels (left). A, Different TE insertion results in reduced and variegated expression of the gene (right). B, Insertion of a DNA TE in the *VvTFL1A* promoter increases branches significantly in cluster structure of grape (a wild-type is shown on the left with the mutant on the right). C, The orange "Navalina" shows limited expression of *Ruby* and the flesh is yellow (upper left). In "Tarocco", an LTR-type retrotransposon drives expression in the flesh with a resultant red color (upper right). Recombination between the two LTRs of the retrotransposon resulted in enhanced expression of Ruby and produced cultivar "Maro I" with purple flesh (lower left). "Jingxian", a Chinese variant with a similar retrotransposon in same allele also exhibits red flesh (lower right). D, Insertion of an LTR-type retrotransposon reduced *Vvmby1A* expression that changes the grape skin from purple (top pannel) to white (lower pannel). E, Ectopic expression of *IQD12* (SUN locus) caused by retrotransposon produces long tomato fruits (a wild-type is shown on the left with the mutant on the right). F, Insertion of a *copia*-like retrotransposon in the upstream region of *tb1* reduced branch formation significantly during the domestication from teosinte (left) to maize (right). G, Transcriptional activation of a rice disease resistance gene *pit* by an LTR-type retrotransposon enhances fungal resistance in resistant species (left) compared to its variant (right).

miniature inverted repeat transposable element (MITE) produces siRNAs to block expression of the flanking genes by H3K9me2 methylation to control agricultural traits in rice (Wei et al., 2014). In figure 2C, the first intron of the *FLOWERING LOCUS C* (*FLC*) gene was inserted with a DNA TE (*mutator-like element, MULE*) in Ler allele, the epigenetic silencing of the TE repress the *FLC* expression by siRNA-involved silencing machinery. Researchers showed that the transcript of the *Athila* LTR-type retrotransposon, through the activity of RNA-dependent RNA polymerase 6 (RDR 6), Dicer-like 2 and 4, was processed into siRNAs in *Arabidopsis* pollens (McCue et al., 2012). One of these TE-siRNAs incorporated into argonaute (AGO) to target *UBP1b* and inhibit its translation (Figure 2D).

In summary, TE may act as *cis*- or *trans*-regulating factors to modulate the expression of flanking or distant genes, respectively. TEs may cause transcriptional gene silencing (TGS) by methylating DNA (McCue et al., 2012) or histone, or impose post-transcriptional gene silencing (PTGS) by DNA cleavage or translational repression. But small RNA (miRNA or siRNA) might be the common actor in these different mechanisms based on recent experimental data (Cui et al., 2014; Yelina et al., 2015; He et al., 2015).

TE EFFECTS ON COTTON GENOME EVOLUTION

Cotton is the largest source of natural textile fiber and a significant oilseed crop. The cotton (*Gossypium* genus) with

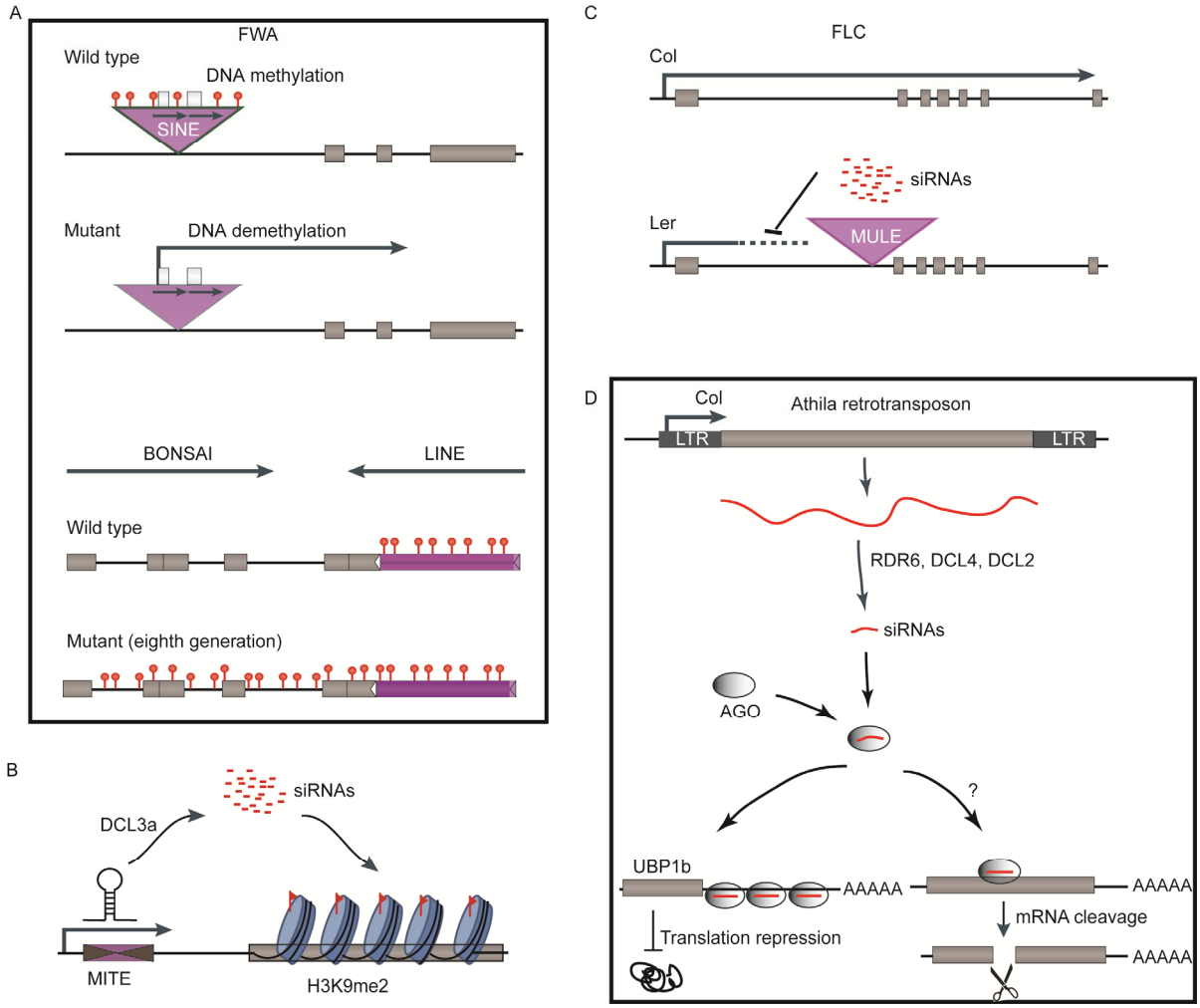

Figure 2 Major types of TE-mediated epigenetic effects on gene regulation. A, Two examples of epigenetic changes in in the *ddm-1* (a chromatin remodeling protein responsible for DNA methylation) background. Upper panel, the methylation loss of a SINE that contains a short tandem duplication (black arrows) in the promoter region of the target gene *FWA* in mutant *ddm1*, would start the *FWA* transcription. Lower panel, the methylation of a LINE spreads its methylation modification to downstream gene *BONSAI* after eight generations in *ddm1* mutant background. B, MITE-derived siRNAs repress the expression of its flanking gene by H3K9me2 methylation. C, A *MULE* insertion degrades target mRNA molecules through siRNA-mediated RNAi pathway. D, LTR-type retrotransposon-derived siRNAs negatively regulate gene expression by translational repression or transcript degradation in a distance-independent manner.

more than 50 species is the largest and most widely distributed genus in the *Gossypieae* tribe (Wendel et al., 2009). A three-fold whole genome duplication (WGD) events in ancestor flowering plants resulted in the production of a genome resembling that of grape (*Vitis vinifera*) about ~115 million years ago (MYA) (Jaillon et al., 2007). The *Gossypieae* tribe experienced another six-fold WGD soon after its branching apart from *T. cacao* (*Theobroma cacao*) about 33 MYA (Jaillon et al., 2007). The *Gossypium* genus separated from Kokia and *Gossypioides* about ~12.5 MYA (Jaillon et al., 2007). The A genome that produces the first spinnable fibers was evolved soon after its divergence from the F genome less than 5 MYA. The allotetraploid (AtAtDtDt) species *G. hirsutum* and *G. barbadense* were formed following an interspecific hybridization between the A-genome diploid and the D-genome diploid 1–2 MYA (Figure 3). As the genus diversified and spread in different habitats, it underwent dramatic genome variation to create adaptive plasticity for natural selection. At present, eight diploid genome groups (A, B, C, D, E, F, G and K) are recognized based on cytogenetic and genetic observations of chromosome pairing behavior, chromosome sizes, and relative fertility in interspecific hybrids (Hawkins et al., 2006).

The genome diversity corresponded fully to the wide-range phenotypic variations in the *Gossypium* genus, such as herbaceous to woody stem, corolla colors (almost span a rainbow), leaf shape and seed size (Fryxell et al., 1979). Fiber or the so-called seed coverings are extraordinarily diverse in *Gossypium* that ranged from nearly glabrous to short and brown hairs, finally to long, fine white fibers in cultured allotetraploid species. Spinnable fiber evolved only in the ancestor of modern A-genome cottons, which afterwards granted the ability to form tetraploid species. Two A-genome diploid species *G. arboreum* and *G. herbaceum*,

and two allotetraploid species, *G. hirsutum* and *G. barbadense* were independently domesticated for cotton fiber traits at least 4,000 years ago (Figure 4) (Dillehay et al., 2007).

TES FOUND IN DIFFERENT COTTON GENOMES

Polyploidization and genome size variation may hold the

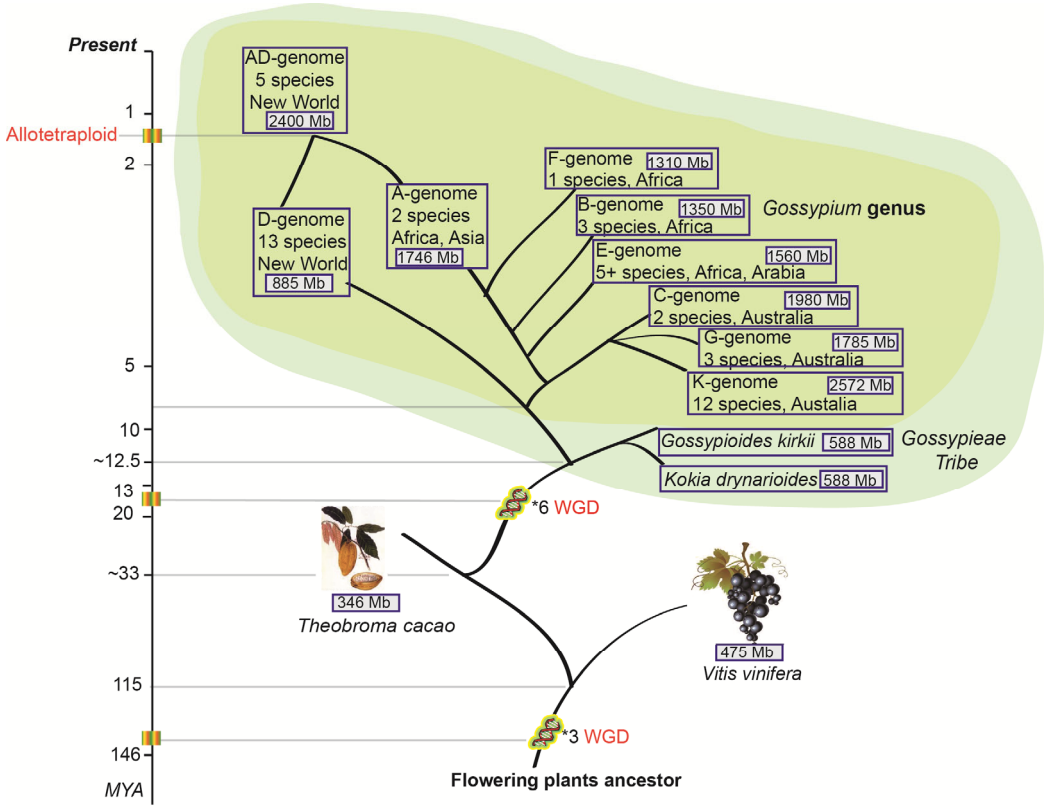

Figure 3 A schematic drawing shows the evolutionary history of the *Gossypium* genus. Polyploidy events are indicated by colored boxes on *Y* axis in log scale that represent the time.

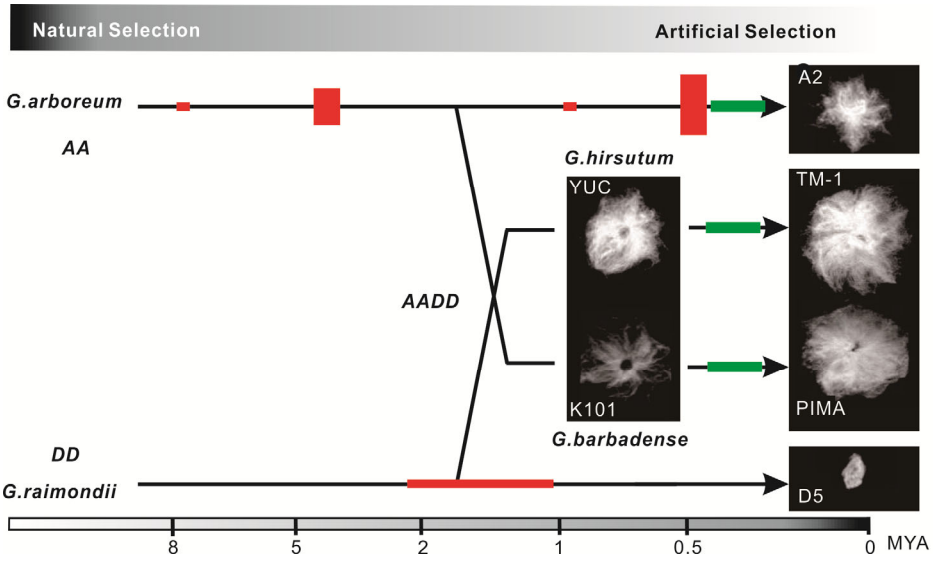

Figure 4 TE expansion during the evolution of diploid and allotetraploid cotton species. Images of a single seed with cotton fiber are shown from *G. arboreum* (AA genome: A_2), *G. raimondii* (DD genome: D_5), a wild form (YUC) and a domesticated species (TM-1) of *G. hirsutum* (AtAtDtDt genome), a wild form (K101) and a domesticated species (PIMA) of *G. barbadense* (AtAtDtDt genome). The red boxes indicate the LTR insertion events, the width and length respectively represent the intensity and duration of TE insertion based on previous studies (Wang et al., 2012; Li et al., 2014). The green boxes indicate the domestication history of cotton that started about 4,000 years ago.

key for future improvement in cotton fiber yield and quality. Since 2012, several joint research groups including our lab-reported the draft genomes of two ancestor diploid cotton, *Gossypium raimondii* (D) (Wang et al., 2012) and *Gossypium arboreum* (A) (Li et al., 2014) together with the cultivated allotetraploid *Gossypium hirsutum* (AD) (Li et al., 2015; Cao, 2015). At about the same time, two other joint groups assembled the same *Gossypium raimondii* (D) (Paterson et al., 2012) and *Gossypium hirsutum* (AD) (Zhang et al., 2015). Table 1 shows the annotations for three genomes based on our joint efforts. The D genome contains 40,976 protein-coding genes with 441.4 Mb, or 57% of the genome as TEs. The genome possessed about 398,000 gypsy-type and about 185,000 *copia*-type retrotransposons that accounted for 78.6% of the TEs. The A genome contains 41,330 protein-coding genes with 1,160 Mb, or 68.5% of the genome as TEs. The genome has 1,086,000 gypsy-type and 186,000 *copia*-type retrotransposons. The AD genome is 2,173 Mb in length with 1,471 Mb, or 67.2% of the genome as TEs. It contains about 1,128,000 gypsy-type and 276,000 *copia*-type retrotransposons. Over 90% of the TEs in either A or the AD genomes are composed of retrotransposons, which is significantly different from that of the D genome.

TE EXPANSION IN COTTON GENOMES

In plant, except for polyploidization, TE amplifications appear to be the major contributor to genome size inflation. For example, the maize genome doubled in size in just a few million years, due most likely to increased TE activity (Sanmiguel et al., 1998). In *Gossypium*, more than three-fold variations of genome size, from ~800 Mb (1C) in the D-genome to ~2,500 Mb (1C) in the K-genome, were found despite of the fact that all diploid species share 13 chromosomes (Figure 3). Previous studies predicted that the variations of genome size of diploid species reflect the volume of copy numbers of repeat DNA sequences (Zhao et al., 1998), especially retrotransposon elements (Hawkins et al., 2006).

Through analyzing the TE divergence rate distribution, we confirmed the expansion of retrotransposons in the *G. raimondii* genome during the last 1–3 million years (Wang et al., 2012). In *G. arboreum*, two major clusters of active LTR-type retrotransposons were found at 0–0.5 and 3.5–4.5 MYA, with two more minor sets occurred around 1.0 and 7.0–8.0 MYA (Figure 4), which indicated that bursts of LTR-type retrotransposon activities may be responsible for the two-fold increase in *G. arboreum* genome compared with *G. raimondii*, similar to previously reported for maize (Swigoňová et al., 2004). Further analysis for syntenic blocks present on chromosome 7 of both *G. arboreum* and *G. raimondii* (3.5 and 1.5 Mb in length, respectively) showed that the protein coding genes in these two blocks were highly collinear whereas the block from *G. arboreum* contains 4,098 TEs and the one from *G. raimondii* has only 1,542. A total of 2,377 *Gorge* elements (one group of gypsy-type retrotransposons) were found in this region from *G. arboreum*, whereas only 324 of them were found in the same region from *G. raimondii* (Li et al., 2014).

In allotetraploid *Gossypium hirsutum,* based on calculation of spontaneous mutation rate and transcriptome data, we suggested that LTR-type retrotransposons were active in allotetraploid genome and *copia* elements were remarkably more active than *gypsy* in the most recent 0–1 million years (Li et al., 2015). Analysis also showed that there were higher proportions of *copia* located near coding genes than *gypsy* and TEs of the Dt subgenome tend to be more active than that of At subgenome after the tetraploidization. Thus we believe that rapid structure and epigenetic reorganization of genome occurred during the early stage of polyploid formation.

TES MAY BE INVOLVED IN COTTON FIBER DEVELOPMENT

Cotton fiber is unicellular, unbranched, simple trichome cell derived from ~25% of the epidermal cells (>25,000 cells) of

Table 1 Comparisons of TE content among *G. raimondii*, *G. arboreum* and *G. hirsutum* genomes[a)]

		G. raimondii			*G. arboreum*			*G. hirsutum*		
		Number	Size (Mb)	% of genome	Number	Size (Mb)	% of genome	Number	Size (Mb)	% of genome
Total scaffolds		4,715	775	100.00	7,914	1,694	100.00	8,591	2,173	100.00
Genes		40,976	116	14.90	41,330	105	6.20	76,943	220	9.50
TEs		–	441	57.09	–	1,146	67.64	–	1,464	67.36
Type I	*LINE*	43,524	12	1.50	33,123	20	1.20	40,543	34	1.56
	SINE	1,924	0.7	0.09	521	0.2	0.01	902	0.7	0.03
	LTR-Gypsy	397,668	262	33.80	1,086,344	945	55.80	1,128,113	1142	52.54
	LTR-Copia	184,815	86	11.10	186,293	93	5.50	276,187	182	8.36
Type II		103,923	16	2.10	74,590	11	0.63	98,115	55	2.52
Unclassified		–	66	8.50	–	76	4.50	–	51	2.35

a) *: Data pooled from references (Fryxell, 1979; Dillehay et al., 2007; Wang et al., 2012).

a developing cottonseed (Li et al., 2005). Great differences in fiber properties are found among the three cotton species. For example, the allotetraploid *G. hirsutum* usually produces fibers with >3 cm in lengths, whereas *G. arboreum* produces fibers of 1.3–1.5 cm long, and no spinnable fiber is produced by *G. raimondii* (Figure 4). The *G. hirsutum* fiber cells undergo fast elongation until ~30 days post anthesis (DPA), whereas those of *G. arboreum* stop growth around 20 DPA. Comparative analysis with that of *T. cacao* and *A. thaliana* revealed that the cotton genomes harbored higher TE content and more TEs were inserted near (within 1 kb of) the cotton genes (Wang et al., 2012; Li et al., 2014; Li et al., 2015). Although the protein coding capacities of these two subgenomes are essentially maintained in the allotetraploid, the expression patterns of a large number of functional genes were significantly different from any of the diploid ancestor. These results indicate that the TE-mediated gene regulation may modulate the unbalanced expression of homologous gene pairs in the allopolyploid.

In the past several decades, many genes involved in regulating fiber elongation were identified. Here we investigated TE distributions in the close vicinity of these identified gene loci. Figure 5A shows the genes that have TE insertion in upstream of coding sequence within the scope of ~5 kb and their roles in cotton fiber development based on previous reports (Huang et al., 2013; Han et al., 2013; Li et al., 2005; Wang et al., 2009; Shi et al., 2006; Qin et al., 2007; Qin et al., 2005; Li et al., 2013; Xu et al., 2009; Shan et al., 2014; Walford et al., 2011; Wang et al., 2013; Luo et al., 2007; Pei, 2015). We identified a *copia*-like retrotransposon insertion in the promoter region of a MYB- domain transcription factor only in the Dt subgenome that is positively correlated with its higher expression in the allotetraploid (Figure 5B). Similarly, a LINE re-

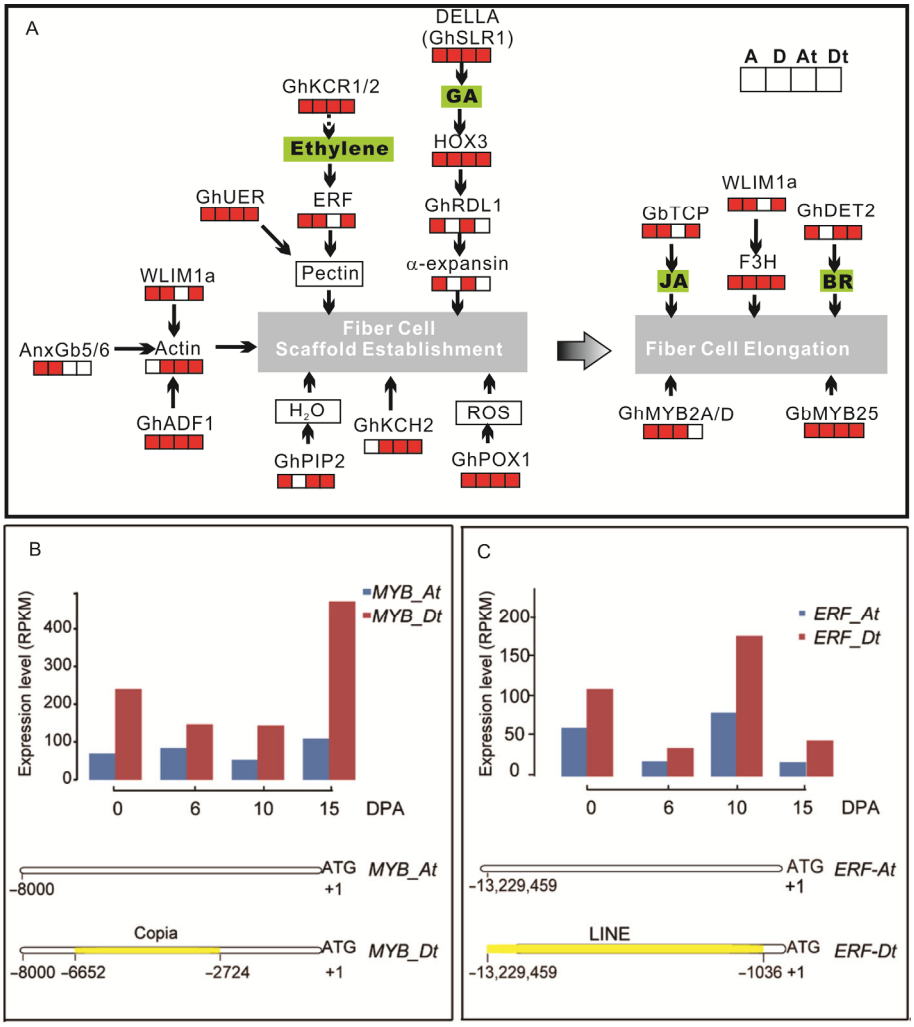

Figure 5 TEs may regulate cotton fiber development. A, Postulated regulatory network during cotton fiber cell elongation and a summary of TE insertion events identified from genomic studies. Different cotton genomes are always arranged as reported in the key shown on top right. Red color indicates the presence of a TE within 5 kb upstream flanking region. A white box indicates that no TE was inserted in the same gene from the particular genome. B, A *copia* inserted in promoter region of one *DtMYB* (upper panel) enhanced its expression significantly during *G. hirsutum* fiber cell development (lower panel). C, A LINE inserted in the *DtERF* promoter region (upper panel) enhanced its expression significantly during *G. hirsutum* fiber cell development (lower panel).

trotransposons insertion in promoter of an *ethylene response factor* (*ERF*) gene in Dt subgenome is related to the Dt expression bias of homologous genes during the ovule cell differentiation process (Figure 5C). Both genes were previously known to be essential for cotton fiber and trichome development (Shi et al., 2006; Qin et al., 2007; Walford et al., 2011; Zhang et al., 2010).

Thus TEs insertion in cotton development associated genes and their polymorphism among AA, DD, Dt and At genomes/subgenomes may be key to understand the vast transcript level and expression pattern variations among different cotton genomes. Also, these data seems to suggest an indispensible role for TEs during the evolution and artificial selection of cotton fiber associated traits.

CONCLUDING REMARKS

The assembled two diploid ancestor cottons, *Gossypium raimondii* (D) and *Gossypium arboreum* (A) together with the cultivated allotetraploid *Gossypium hirsutum* (AD) genomes provide references for genome re-sequencing in multiple species of *Gossypium*. Due to a recent burst of high throughput sequencing based technologies, identification of TE polymorphisms by genome re-sequencing gradually became a reality. Transcriptomic, epigenomic analyses and especially functional studies such as gene editing may be used to elucidate TE-mediated phenotypic changes in *Gossypium*. The genetic resources obtained from these researches may enhance cotton production and improve fiber quality in the near future. We hope that the discovery in molecular mechanisms from different *Gossypium* genomes will benefit studies of other plants, and even animals.

This work was supported by the National Natural Science Foundation of China (90717009), and the Chinese National Basic Research Program of the Ministry of Science and Technology of China (2010CB126000).

Bennetzen, J.L. (2007). Patterns in grass genome evolution. Curr Opin Plant Biol 10, 176–181.

Bennetzen, J.L., and Wang, H. (2014). The contributions of transposable elements to the structure, function, and evolution of plant genomes. Annu Rev Plant Biol 65, 505–530.

Butelli, E., Licciardello, C., Zhang, Y., Liu, J., Mackay, S., Bailey, P., Reforgiato-Recupero, G., Martin, C. (2012). Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. Plant Cell 24, 1242–1255.

Cao, X. (2015). Whole genome sequencing of cotton—a new chapter in cotton genomics. Sci China Life Sci 58, 515–516.

Cui, X., and Cao, X. (2014). Epigenetic regulation and functional exaptation of transposable elements in higher plants. Curr Opin Plant Biol 21, 83–88.

Dillehay, T.D., Rossen, J., Andres, T.C., and Williams, D.E. (2007). Preceramic adoption of peanut, squash, and cotton in northern Peru. Science 316, 1890–1893.

Fernandez, L., Torregrosa, L., Segura, V., Bouquet, A., and Martinez-Zapater, J.M. (2010). Transposon-induced gene activation as a mechanism generating cluster shape somatic variation in grapevine. Plant J 61, 545–557.

Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks. Nat Rev Genet 9, 397–405.

Fryxell, P.A. (1979). Natural history of the cotton tribe. Texas: Texas A&M University Press.

McClintock, B. (1948). Mutable loci in maize. Carnegie Inst Wash Yearb 47, 155–169.

McClintock, B. (1956). Controlling elements and the gene. Cold Spring Harb Sym 21, 197–216.

Han, L., Li, Y., Wang, H., Wu, X., Li, C., Luo, M., Wu, S., Kong, Z., Pei, Y., Jiao, G., and Xia, G. (2013). The dual functions of WLIM1a in cell elongation and secondary wall formation in developing cotton fibers. Plant Cell 25, 4421–4438.

Hawkins, J.S., Kim, H., 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 Res 16, 1252–1261.

Hayashi, K., and Yoshida, H. (2009). Refunctionalization of the ancient rice blast disease resistance gene pit by the recruitment of a retrotransposon as a promoter. Plant J 57, 413–425.

He, H., Yang, T., Wu, W., and Zheng, B. (2015). Small RNAs in pollen. Sci China Life Sci 58, 246–252.

Huang, Y., Wang, J., Zhang, L., and Zuo, K. (2013). A cotton annexin protein AnxGb6 regulates fiber elongation through its interaction with actin 1. PLoS One 8, e66160.

Jaillon, O., Aury, J.M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., Vezzi, A., Legeai, F., Hugueney, P., Dasilva, C., Horner, D., Mica, E., Jublot, D., Poulain, J., Bruyere, C., Billault, A., Segurens, B., Gouyvenoux, M., Ugarte, E., Cattonaro, F., Anthouard, V., Vico, V., Del Fabbro, C., Alaux, M., Di Gaspero, G., Dumas, V., Felice, N., Paillard, S., Juman, I., Moroldo, M., Scalabrin, S., Canaguier, A., Le Clainche, I., Malacrida, G., Durand, E., Pesole, G., Laucou, V., Chatelet, P., Merdinoglu, D., Delledonne, M., Pezzotti, M., Lecharny, A., Scarpelli, C., Artiguenave, F., Pe, M.E., Valle, G., Morgante, M., Caboche, M., Adam-Blondon, A.F., Weissenbach, J., Quetier, F., Wincker, P., and French-Italian, P. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449, 463–465.

Kim, H.J., and Triplett, B.A. (2001). Cotton fiber growth *in planta* and *in vitro*. Models for plant cell elongation and cell wall biogenesis. Plant Physiol 127, 1361–1366.

Kinoshita, Y., Saze, H., Kinoshita, T., Miura, A., Soppe, W.J.J., Koornneef, M., and Kakutani, T. (2007). Control of *FWA* gene silencing in *Arabidopsis thaliana* by sine-related direct repeats. Plant J 49, 38–45.

Kobayashi, S., Goto-Yamamoto, N., and Hirochika, H. (2004). Retrotransposon-induced mutations in grape skin color. Science 304, 982–982.

Li, D., Ruan, X., Zhang, J., Wu, Y., Wang, X., and Li, X. (2013). Cotton plasma membrane intrinsic protein 2s (PIP2s) selectively interact to regulate their water channel activities and are required for fibre development. New Phytol 199, 695–707.

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., Zhu, Y., Wang, J., and Yu, S. (2014). Genome sequence of the cultivated cotton *Gossypium arboreum*. Nat Genet 46, 567–572.

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. Nat Biotechnol 33, 524–530.

Lisch, D. (2013). How important are transposons for plant evolution? Nat

Rev Genet 14, 49–61.
Li, X., Fan, X., Wang, X., Cai, L., and Yang, W. (2005). The cotton *ACTIN1* gene is functionally expressed in fibers and participates in fiber elongation. Plant Cell 17, 859–875.
Li, Y., and Dooner, H.K. (2009). Excision of helitron transposons in maize. Genetics 182, 399–402.
Luo, M., Xiao, Y., Li, X., Lu, X., Deng, W., Li, D., Hou, L., Hu, M., Li, Y., and Pei, Y. (2007). GhDET2, a steroid 5 alpha-reductase, plays an important role in cotton fiber cell initiation and elongation. Plant J 51, 419–430.
McCue, A.D., Nuthikattu, S., Reeder, S.H., and Slotkin, R.K. (2012). Gene expression and stress response mediated by the epigenetic regulation of a transposable element small RNA. PLoS Genet 8, e1002474.
Oliver, K.R., McComb, J.A., and Greene, W.K. (2013). Transposable elements: powerful contributors to angiosperm evolution and diversity. Genome Biol Evol 5, 1886–1901.
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, Mu., 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., Van Deynze, A., 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, 423–427.
Pei, Y. (2015). The homeodomain—containing transcription factor, GhHOX3, is a key regulator of cotton fiker elongation. Sci China Life Sci 58, 309–310.
Piegu, B., Guyot, R., Picault, N., Roulin, A., Saniyal, A., Kim, H., Collura, K., Brar, D.S., Jackson, S., Wing, R.A., and Panaud, O. (2006). Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in oryza australiensis, a wild relative of rice. Genome Res 16, 1262–1269.
Qin, Y., Hu, C., Pang, Y., Kastaniotis, A.J., Hiltunen, J.K., and Zhu, Y. (2007). Saturated very-long-chain fatty acids promote cotton fiber and arabidopsis cell elongation by activating ethylene biosynthesis. Plant Cell 19, 3692–3704.
Qin, Y., Pujol, F.M.A., Shi, Y., Feng, J., Liu, Y., Kastaniotis, A.J., Hiltunen, J.K., and Zhu, Y. (2005). Cloning and functional characterization of two cDNAs encoding NADPH-dependent 3-ketoacyl-CoA reductased from developing cotton fibers. Cell Res 15, 465–473.
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. Ann Bot 82 (Supplement A), 37–44.
Saze, H., and Kakutani, T. (2007). Heritable epigenetic mutation of a transposon-flanked *Arabidopsis* gene due to lack of the chromatin-remodeling factor DDM1. EMBO J 26, 3641.
Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A., Minx, P., Reily, A.D., Courtney, L., Kruchowski, S.S., Tomlinson, C., Strong, C., Delehaunty, K., Fronick, C., Courtney, B., Rock, S.M., Belter, E., Du, F.Y., Kim, K., Abbott, R.M., Cotton, M., Levy, A., Marchetto, P., Ochoa, K., Jackson, S.M., Gillam, B., Chen, W., Yan, L., Higginbotham, J., Cardenas, M., Waligorski, J., Applebaum, E., Phelps, L., Falcone, J., Kanchi, K., Thane, T., Scimone, A., Thane, N., Henke, J., Wang, T., Ruppert, J., Shah, N., Rotter, K., Hodges, J., Ingenthron, E., Cordes, M., Kohlberg, S., Sgro, J., Delgado, B., Mead, K., Chinwalla, A., Leonard, S., Crouse, K., Collura, K., Kudrna, D., Currie, J., He, R.F., Angelova, A., Rajasekar, S., Mueller, T., Lomeli, R., Scara, G., Ko, A., Delaney, K., Wissotski, M., Lopez, G., Campos, D., Braidotti, M., Ashley, E., Golser, W., Kim, H., Lee, S., Lin, J.K., Dujmic, Z., Kim, W., Talag, J., Zuccolo, A., Fan, C., Sebastian, A., Kramer, M., Spiegel, L., Nascimento, L., Zutavern, T., Miller, B., Ambroise, C., Muller, S., Spooner, W., Narechania, A., Ren, L.Y., Wei, S., Kumari, S., Faga, B., Levy, M.J., McMahan, L., Van Buren, P., Vaughn, M.W., Ying, K., Yeh, C.T., Emrich, S.J., Jia, Y., Kalyanaraman, A., Hsia, A.P., Barbazuk, W.B., Baucom, R.S., Brutnell, T.P., Carpita, N.C., Chaparro, C., Chia, J.M., Deragon, J.M., Estill, J.C., Fu, Y., Jeddeloh, J.A., Han, Y.J., Lee, H., Li, P., Lisch, D.R., Liu, S., Liu, Z., Nagel, D.H., McCann, M.C., SanMiguel, P., Myers, A.M., Nettleton, D., Nguyen, J., Penning, B.W., Ponnala, L., Schneider, K.L., Schwartz, D.C., Sharma, A., Soderlund, C., Springer, N.M., Sun, Q., Wang, H., Waterman, M., Westerman, R., Wolfgruber, T.K., Yang, L., Yu, Y., Zhang, L., Zhou, S., Zhu, Q., Bennetzen, J.L., Dawe, R.K., Jiang, J., Jiang, N., Presting, G.G., Wessler, S.R., Aluru, S., Martienssen, R.A., Clifton, S.W., McCombie, W.R., Wing, R.A., and Wilson, R.K. (2009). The b73 maize genome: complexity, diversity, and dynamics. Science 326, 1112–1115.
Selinger, D.A., and Chandler, V.L. (2001). B-bolivia, an allele of the maize *b1* gene with variable expression, contains a high copy retrotransposon-related sequence immediately upstream. Plant Physiol 125, 1363–1379.
Shan, C., Shangguan, X., Zhao, B., Zhang, X., Chao, L., Yang, C., Wang, L., Zhu, H., Zeng, Y., Guo, W., Zhou, B., Hu, G., Guan, X., Chen, Z., Wendel, J.F., Zhang, T., and Chen, X. (2014). Control of cotton fibre elongation by a homeodomain transcription factor GhHOX3. Nat Commun 5, 5519.
Shi, Y., Zhu, S., Mao, X., Feng, J., Qin, Y., Zhang, L., Cheng, J., Wei, L., Wang, Z., and Zhu, Y. (2006). Transcriptome profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber cell elongation. Plant Cell 18, 651–664.
Studer, A., Zhao, Q., Ross-Ibarra, J., and Doebley, J. (2011). Identification of a functional transposon insertion in the maize domestication gene *tb1*. Nat Genet 43, 1160–1163.
Swigoňová, Z., Lai, J., Ma, J., Ramakrishna, W., Llaca, V., Bennetzen, J.L., and Messing, J. (2004). Close split of sorghum and maize genome progenitors. Genome Res 14, 1916–1923.
Tenaillon, M.I., Hollister, J.D., and Gaut, B.S. (2010). A triptych of the evolution of plant transposable elements. Trends Plant Sci 15, 471–478.
Walford, S.A., Wu, Y., Llewellyn, D.J., and Dennis, E.S. (2011). GhMYB25-like: a key factor in early cotton fibre development. Plant J 65, 785–797.
Wang, H., Wang, J., Gao, P., Jiao, G., Zhao, P., Li, Y., Wang, G., and Xia, G. (2009). Down-regulation of *GhADF1* gene expression affects cotton fibre properties. Plant Biotechnol J 7, 13–23.
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., Wang, J., and Yu, S. (2012). The draft genome of a diploid cotton *Gossypium raimondii*. Nat Genet 44, 1098–1103.
Wang, M., Zhao, P., Cheng, H., Han, L., Wu, X., Gao, P., Wang, H., Yang, C., Zhong, N., Zuo, J., and Xia, G. (2013). The cotton transcription factor TCP14 functions in auxin-mediated epidermal cell differentiation and elongation. Plant Physiol 162, 1669–1680.
Wei, L., Gu, L., Song, X., Cui, X., Lu, Z., Zhou, M., Wang, L., Hu, F., Zhai, J., Meyers, B.C., and Cao, X. (2014). Dicer-like 3 produces transposable element-associated 24-nt siRNAs that control agricultural traits in rice. Proc Natl Acad Sci USA 111, 3877–3882.
Wendel, J., Brubaker, C., Alvarez, I., Cronn, R., and Stewart, J. (2009). Evolution and natural history of the cotton genus. Genet Genomics Cotton 3, 3–22.
Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., and Schulman, A.H. (2007). A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8, 973–982.
Xiao, H., Jiang, N., Schaffner, E., Stockinger, E.J., and van der Knaap, E. (2008). A retrotransposon-mediated gene duplication underlies

morphological variation of tomato fruit. Science 319, 1527–1530.
Xu, T., Qu, Z., Yang, X., Qin, X., Xiong, J., Wang, Y., Ren, D., and Liu, G. (2009). A cotton kinesin GhKCH2 interacts with both microtubules and microfilaments. Biochem J 421, 171–180.
Yang, Q., Li, Z., Li, W., Ku, L., Wang, C., Ye, J., Li, K., Yang, N., Li, Y., Zhong, T., Li, J., Chen, Y., Yan, J., Yang, X., and Xu, M. (2013). CACTA-like transposable element in ZmCCT attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. Proc Natl Acad Sci USA 110, 16969–16974.
Yelina, N., Diaz, P., Lambing, C., and Henderson, I.R. (2015). Epigenetic control of meiotic recombination in plants. Sci China Life Sci 58, 223–231.
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. J Exp Bot 61, 3599–3613.
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. (2015). Sequencing of allotetraploid cotton (*Gossypium hirsutum* l. Acc. Tm-1) provides a resource for fiber improvement. Nat Biotechnol 33, 531–537.
Zhao, X., Si, Y., Hanson, R.E., Crane, C.F., Price, H.J., Stelly, D.M., Wendel, J.F., and Paterson, A.H. (1998). Dispersed repetitive DNA has spread to new genomes since polyploid formation in cotton. Genome Res 8, 479–492.