THE CROP JOURNAL XX (2016) XXX–XXX

HOSTED BY

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Asymmetric evolution and domestication in allotetraploid cotton (*Gossypium hirsutum* L.)

*Lei Fang, Xueying Guan, Tianzhen Zhang**

State Key Laboratory of Crop Genetics and Germplasm Enhancement, Cotton Hybrid R&D Engineering Center (the Ministry of Education), Nanjing Agricultural University, Nanjing 210095, China

ARTICLE INFO

Article history:
Received 10 June 2016
Received in revised form 3 July 2016
Accepted 11 July 2016
Available online xxxx

Keywords:
Polyploidy
Allotetraploid cotton
Asymmetric evolution
Asymmetric domestication

ABSTRACT

Polyploidy plays a major role in genome evolution, which corresponds to environmental changes over millions of years. The mechanisms of genome evolution, particularly during the process of domestication, are of broad interest in the fields of plant science and crop breeding. Upland cotton is derived from the hybridization and polyploidization of its ancient A and D diploid ancestors. As a result, cotton is a model for polyploid genome evolution and crop domestication. To explore the genomic mysteries of allopolyploid cotton, we investigated asymmetric evolution and domestication in the A and D subgenomes. Interestingly, more structural rearrangements have been characterized in the A subgenome than in the D subgenome. Correspondingly, more transposable elements, a greater number of lost and disrupted genes, and faster evolution have been identified in the A subgenome. In contrast, the centromeric retroelement (RT-domain related) sequence of tetraploid cotton derived from the D subgenome progenitor was found to have invaded the A subgenome centromeres after allotetrapolyploid formation. Although there is no genome-wide expression bias between the subgenomes, as with expression-level alterations, gene expression bias of homoeologous gene pairs is widespread and varies from tissue to tissue. Further, there are more positively selected genes for fiber yield and quality in the A subgenome and more for stress tolerance in the D subgenome, indicating asymmetric domestication. This review highlights the asymmetric subgenomic evolution and domestication of allotetraploid cotton, providing valuable genomic resources for cotton research and enhancing our understanding of the basis of many other allopolyploids.

1. Introduction

Polyploidy, or whole-genome duplication (WGD), plays a prominent role in shaping genome evolution and genetic diversity in many plants and some animals [1–4]. As the addition of a complete set of chromosomes to a genome, polyploidization represents one of the most dramatic mutations known to occur and can affect cell size, body size, genomic stability, gene expression, and evolution rate [4]. Allopolyploidy has long been recognized as an important mode of plant speciation, and results from hybridization and genome doubling. However, increased gene and genome dosages in allopolyploids often cause genomic instabilities, chromosome imbalances, regulatory incompatibilities, and reproductive failure. As a result, changes in DNA sequence, chromatin modifications, and expression of homoeologous

* *Corresponding author*. Fax: +86 25 84395307.
E-mail address: cotton@njau.edu.cn (T. Zhang).
Peer review under responsibility of Crop Science Society of China and Institute of Crop Science, CAAS.

http://dx.doi.org/10.1016/j.cj.2016.07.001

Please cite this article as: L. Fang, et al., Asymmetric evolution and domestication in allotetraploid cotton (*Gossypium hirsutum* L.), The Crop Journal (2016), http://dx.doi.org/10.1016/j.cj.2016.07.001

genes are necessary for establishing a compatible genome and facilitating adaptive evolution [1].

Cotton production provides income for approximately 100 million families, and approximately 150 countries are involved in cotton import and export. Its economic impact is estimated to be approximately $500 billion per year worldwide [5]. Cotton is a model system for studying cell fate determination, cellulose biosynthesis, and the effects of polyploidy on domestication. *Gossypium* includes five allotetraploid ($2n = 4\times = 52$) and ~45 diploid ($2n = 2\times = 26$) species [6]. The New World allotetraploids *G. hirsutum* L. $(AD)_1$ and *G. barbadense* L. $(AD)_2$ and the Old World diploids *G. herbaceum* L. (A_1) and *G. arboreum* L. (A_2) were independently domesticated for fiber production. A- and D-genome species diverged from a common ancestor approximately 5–10 million years ago (MYA), and polyploidization of the closest extant progenitors, *G. herbaceum* L. (A_1) and *G. raimondii* Ulbrich (D_5) occurred 1–2 MYA [7]. Interestingly, the A genome (1700 megabases, Mb) is twice the size of the D genome (885 Mb) [8–10]. Over 95% of the cultivated cotton crop in the world is Upland cotton (*G. hirsutum*) (http://www.cotton.org/2012), which has longer, stronger and finer fibers along with a dramatically enhanced fiber yield relative to the progenitor species *G. arboretum* and *G. raimondii*. These increases likely result from polyploidization and domestication.

Genome sequences provide many insights into plant genome evolution, suggesting a history of repeated, episodic WGD events and the ubiquity of variation in genome size across plant species. Polyploidy generates new genomic interactions, "genomic and transcriptomic shock", resulting in a new polyploid lineage [3]. Since the *Arabidopsis thaliana* genome was published 14 years ago [11], more than 55 plant genomes representing 49 different species have been sequenced [12], including recent paleopolyploids such as maize and soybean [13,14], hexaploid wheat [15], and the allopolyploid *Brassica napus*. Rapid genomic reorganization and massive gene loss have been observed in paleopolyploid maize [16], newly allopolyploid wheat [17] and *Brassica* [18], but only small changes are present in *Arabidopsis* [19] and a relatively stable genome organization has been verified in cultivated hexaploid wheat [20] and *Brassica napus* [21]. These genomic changes may ultimately promote adaptive speciation. Polyploidy manifests at the chromosomal, physiological, and organismal level through alterations in genome structure, gene content, and expression [3].

To date, the whole-genome sequence of allotetraploid cotton [22,23] and its corresponding progenitor species [9,10] have been completed. The high-quality genome of allotetraploid Upland cotton [22] provides an efficient tool for systematically exploring the genomic mysteries of polyploidy. Our purpose in this short review is to highlight the asymmetric subgenomic evolution and domestication of allotetraploid cotton, with a focus on genomic structural rearrangements, gene loss and disruption, evolution rate, centromeric region transfer, expression bias, and positive selection in the A and D subgenomes (Fig. 1).

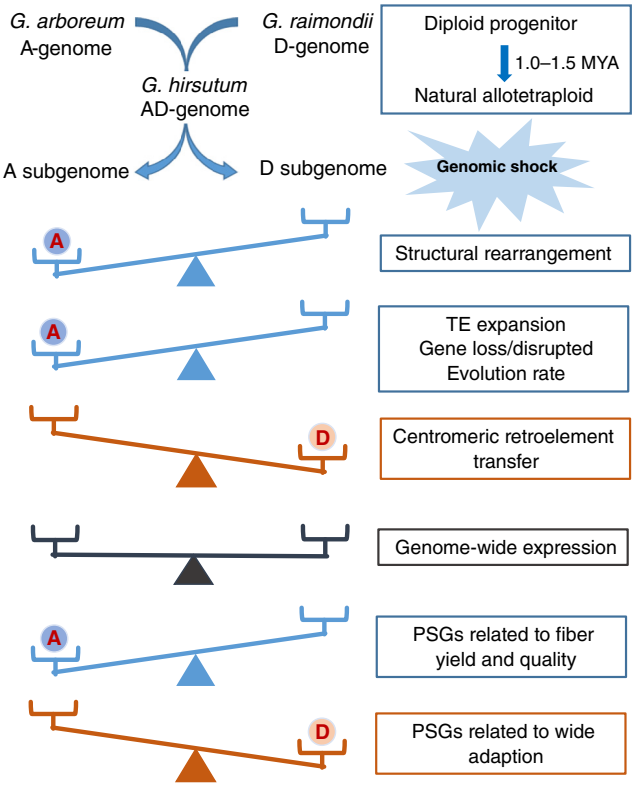

Fig. 1 – Asymmetric genome and gene evolution in allotetraploid cotton. Allotetraploid cotton contains A and D subgenomes, and was formed approximately 1–1.5 MYA from an A-genome and a D-genome diploid. In the polyploidy event, responses to genome doubling included genomic alterations (structural arrangement, TE expansion, lost and disrupted genes, evolution rate, centrometric retroelement transfer, and positively selected genes (PSGs) and transcriptomic alterations. The weighting scale biased to A (blue) or D (orange) subgenomes indicates the corresponding asymmetric evolution event shown in the right box.

2. Asymmetric structural rearrangements of subgenomes

Allopolyploid plants often show major changes in genome structure and function induced by the reconciliation of two sets of diverged genomes and their regulatory interactions [24], including inter-genomic chromosomal rearrangements, gene loss, gene conversion, divergence, and functional diversification of duplicated genes [25]. Allotetraploid cottons contain A and D subgenomes inherited from their original A and D diploid genomes. Comparative genome analysis of cotton subgenomes and extant diploids shows that the high degrees of colinearity and synteny between the A and D subgenomes and the D-progenitor genome are retained after polyploidization, as reported in hexaploid wheat [20]. However, across the genomes, there are some instances of structural rearrangement. A total of nine translocations and 28 inversions have been identified [22,26], including 19 and 18 rearrangements in the A and D

Please cite this article as: L. Fang, et al., Asymmetric evolution and domestication in allotetraploid cotton (*Gossypium hirsutum* L.), The Crop Journal (2016), http://dx.doi.org/10.1016/j.cj.2016.07.001

Table 1 – Genome structural variations of the A and D subgenome of TM-1 identified by comparison with the D genome *G. raimondii* [10].

A subgenome					D subgenome				
A subgenome vs. D genome	Physical interval (Mb)	Block (Mb)	Chr. length (Mb)	% of Chr.	D subgenome vs. D genome	Physical interval (Mb)	Block (Mb)	Chr. Length (Mb)	% of Chr.
^{I}A01-Gr_Chr02	77.55–84.4	6.85	99.88	6.86	^{I}D01- Gr_Chr02	38.6–40.47	1.87	61.46	3.04
^{I}A02-Gr_Chr03	38.03–53.01	14.98	83.45	17.95	^{I}D02- Gr_Chr05	29.83–30.85	1.02	67.28	1.52
^{I}A02-Gr_Chr03	60.09–65.93	5.84	83.45	7.00	^{I}D02- Gr_Chr05	47.13–48.82	1.69	67.28	2.51
^{I} A02-Gr_Chr03	73.29–78.29	5.00	83.45	5.99	^{I}D02- Gr_Chr05	33.50–39.68	6.18	67.28	9.18
^{I}A03-Gr_Chr05	36.98–43.98	7.00	100.26	6.98	^{I}D03- Gr_Chr03	17.57–29.46	11.89	46.69	25.47
^{I}A03-Gr_Chr03	14.51–21.50	6.99	100.26	6.97	^{I}D05- Gr_Chr09	56.74–58.17	1.43	61.93	2.31
^{I}A05-Gr_Chr12	67.01–76.01	9.00	92.05	9.78	^{I}D09- Gr_Chr06	20.24–26.37	6.13	51.00	12.01
^{I}A09-Gr_Chr09	30.54–43.80	13.26	75.00	17.68	^{I}D10- Gr_Chr11	40.84–41.84	1.00	63.37	1.58
^{I}A10-Gr_Chr11	9.26–12.83	3.57	100.87	3.54	^{I}D10- Gr_Chr11	36.86–38.73	1.88	63.37	2.96
^{I}A10-Gr_Chr11	13.99–20.71	6.72	100.87	6.66	^{I}D11- Gr_Chr07	39.09–40.86	1.77	66.09	2.68
^{I}A10-Gr_Chr11	20.89–75.50	54.61	100.87	54.14	^{I}D11- Gr_Chr07	31.65–33.03	1.37	66.09	2.08
^{I}A11-Gr_Chr07	37.76–57.00	19.24	93.32	20.62	^{I}D12- Gr_Chr08	14.35–28.36	14.01	59.11	23.70
^{I}A12-Gr_Chr08	4.49–7.73	3.24	87.48	3.70	^{I}D13- Gr_Chr13	50.55–52.94	2.39	60.53	3.96
^{I}A12-Gr_Chr08	39.16–52.16	13.00	87.48	14.86	^{T}D03- Gr_Chr13	12.94–21.51	8.57	46.69	18.36
^{I}A13-Gr_Chr13	6.55–15.45	8.90	79.96	11.13	^{T}D04- Gr_Chr09	23.94–37.59	13.65	51.45	26.52
^{RT}A02-Gr_Chr03	38.03–83.45	45.42	83.45	54.43	^{T}D11- Gr_Chr03	41.10–43.07	1.97	66.09	2.98
^{RT}A03-Gr_Chr05	22.56–100.26	77.70	100.26	77.50	^{T}D13- Gr_Chr10	20.76–21.98	1.22	60.53	2.01
^{RT}A04-Gr_Chr09	0–46.26	46.26	62.91	73.53	^{T}D13- Gr_Chr01	22.07–26.66	4.59	60.53	7.58
^{RT}A05-Gr_Chr12	67.01–92.05	25.04	92.05	27.20					
Mean		19.61			Mean		4.59		
Median		9.00			Median		1.93		
Total		372.60			Total		82.6		

These structural variations were described in previous study [22]. Chr chromosome; ^{I} inversion; ^{T} translocation; ^{RT} reciprocal translocation.

subgenomes, respectively (Table 1). Among them are two postpolyploidization reciprocal translocations between A02 and A03 and between A04 and A05 in the A-subgenome chromosomes. Additional translocations and inversions are located mainly in A09, A13, D03, D04, and D13. Although the numbers of structural rearrangements in the A and D subgenomes are similar, the length of rearrangements was significantly greater in the A than in the D subgenome (372.6 vs. 82.6 Mb) ($P < 0.01$, Fisher's exact test). The average length was 19.6 Mb in the A subgenome, a length that was also significantly greater than the 4.6 Mb in the D subgenome ($P < 0.01$, Fisher's exact test). By comparison of the genetic map comprising 4,999,048 SNP loci distributed unevenly in 26 allotetraploid cotton linkage groups and the corresponding physical molecules of the D genome sequence of *G. raimondii* [10], a total of 34 structural rearrangements were also identified, including 23 in the A and 11 in the D subgenome (Table S1) [26]. This finding further supported the above two reciprocal translocations based on comparison between the genetic and physical map. These complex chromosomal rearrangements (inversions and translocations) are considered important factors in cotton evolution. The potential molecular basis of these phenomena merits further study.

3. Asymmetric evolution of the A and D subgenomes

In addition to WGD events, variation in plant genome size may be due to the dynamics of transposable element (TE) proliferation and clearance [27,28]. In *Brassica napus*, the assembled genome contains 34.80% TEs, with an asymmetric distribution in the A_n and C_n subgenomes [21]. Both retro- (22.13%) and DNA (16.67%) TEs appear to be amplified to a greater extent in *B. oleracea* (C_n) relative to *B. rapa* (A_n) (9.43% and 12.04%, respectively) [29]. In allotetraploid cotton, the A subgenome is twice the size of the D subgenome. Relative to the D subgenome (746.8 Mb), the A subgenome (1220.6 Mb) has expanded by 473.8 Mb, resulting from the abundance of TEs (64.80% of the *G. hirsutum* genome) in the A subgenome (843.5 Mb), compared to the D subgenome (433 Mb) [22]. Among these TEs, Gypsy-like elements were the most divergent between the A (362 Mb) and D (136.1 Mb) subgenomes, contributing 47.7% (225.9 Mb) of the increase in the A subgenome. The expansion of TEs in the A and D homoeologous genomes might represent a divergent distribution in their progenitor genomes. However, the TE types and proportions in the subgenomes and their corresponding progenitor genomes are similar, suggesting that most TEs expanded before the formation of allopolyploid cotton and a little expansion occurred in the D subgenome only post-polyploidization [22]. TEs have been implicated as one of the most important factors in gene regulation, gene content, and environmental adaptation [30].

In addition to the asymmetric expansion of TEs, collinearity analysis was performed only between subgenomes and the D-progenitor genome [10], because numerous examples of mis-assembly were present in either the A-progenitor genome [9] or another version of the D-progenitor genome [31]. As a result, a rare gene loss in the allotetraploid cotton was found, including 228 genes in the A subgenome and 141 genes in the D subgenome ($P < 0.01$, Fisher's exact test), suggesting that genes originating in the D genome are more resistant to gene loss [22].

Please cite this article as: L. Fang, et al., Asymmetric evolution and domestication in allotetraploid cotton (*Gossypium hirsutum* L.), The Crop Journal (2016), http://dx.doi.org/10.1016/j.cj.2016.07.001

Similarly, more disrupted genes have been identified in the A subgenome (2425) than in the D subgenome (1887) ($P < 0.01$, Fisher's exact test). These genes contain frame shifts or premature stop codons compared to their orthologous genes and are poorly expressed. These data suggest that the D subgenome is more resistant to gene loss and pseudogenization [22]. Biased gene loss and biased fractionation have been detected in other monocot and eudicot plants [32,33], and very little is known about these phenomena. The biased gene loss is hypothesized to have arisen through alterations in gene expression and local TEs in cotton [34].

Another intriguing finding concerns the asymmetric evolution rates of the subgenomes. Comparison of synonymous substitution values of 21,618 one-to-one orthologous gene sets in the A genome, D genome and two subgenomes, led to an estimate of 6.0–6.3 MYA for the divergence time between the A and D genomes, and the A and D subgenomes were united 1.0–1.5 MYA (with a K_s peak at 0.005 and 0.008). The K_a/K_s value of the A subgenome is higher than that of the D subgenome, as confirmed by Tajima's relative rate [22]. These results suggested that the A subgenome evolved faster than the D subgenome. Indeed, the physically larger A subgenome has been altered more extensively than the D subgenome, indicating an asymmetric evolution even during allotetraploid cotton formation.

4. Intergenomic centromeric retroelement transfer

The centromere is an essential chromosomal region that is difficult to map, clone, and sequence, owing to its highly repetitive sequence and low recombination rate [35]. In recent years, long terminal repeats (LTRs) have been reported in centromeric regions. These LTRs, named GhCR and CRG, have been localized in all 52 chromosomes in allotetraploid cotton and D-genome cotton species and have been validated by fluorescence in situ hybridization (FISH) experiments [36,37]. Based on genome-wide GhCR LTR homology screening, 3515 LTRs are thought to be located in the D subgenome (about 270 per chromosome); a number significantly higher than the 430 in the A subgenome (about 33 per chromosome) [26], and this finding was supported by our FISH hybridization study, showing much weaker signals in the A than in the D subgenome [37]. Although 17 CRGs and 119 GhCRs have been identified in the diploid A genome of *G. arboreum*, there are no reverse transcriptase RT domains from GhCRs LTR retrotransposon. However, one GhCR (RT) domain is located in the A subgenome. In previous studies, GhCR and GhCRG have been verified in allotetraploid and D genome diploid species, but not in A genome diploid species such as *G. arboreum* and *G. herbaceum* [36,37]. These results suggest that some centromeric retrotransposons in the A subgenome were derived from the D-subgenome progenitor of allotetraploid cotton [26].

5. Biased gene expression in subgenomes

It is unclear whether polyploidy and genome doubling has had a large impact on genome-wide alterations in gene expression. Expression of homoeologous genes in allopolyploids could be biased toward one subgenome over another. Biased expression levels could vary in different plant organs, with one third of all genes being biased toward one homoeolog or the other in at least one organ [38]. In cotton, the comprehensive genome-wide RNA-seq analysis of 25,358 homoeologous gene pairs in diverse tissues, and the average number of expressing tissues, showed no evidence of genome-wide expression bias between the two subgenomes ($P > 0.5$, Wilcoxon rank sum test) [22]. This pattern concurs

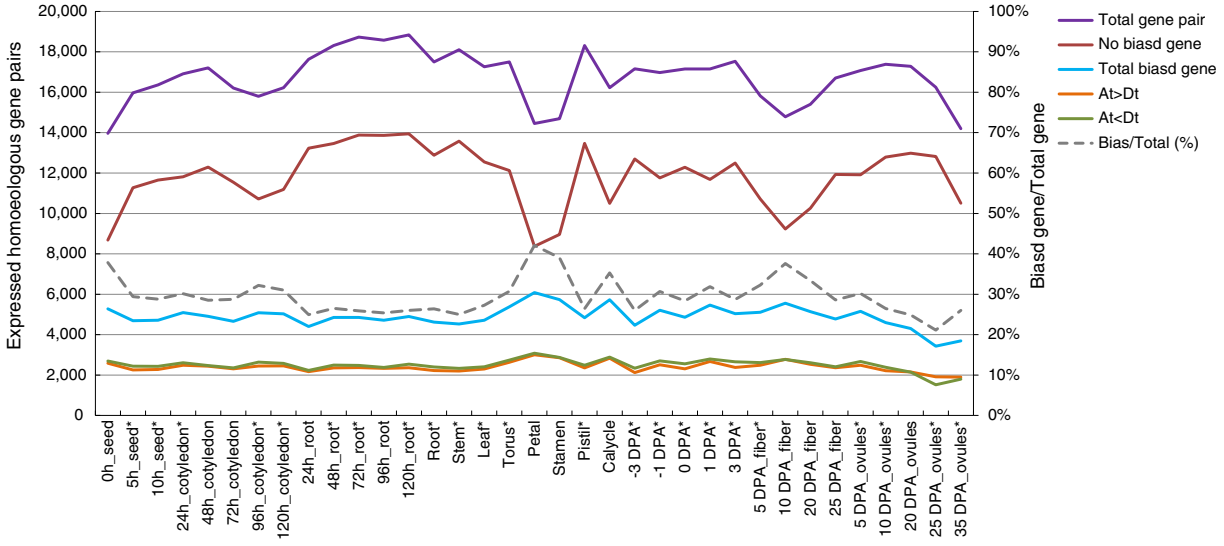

Fig. 2 – Summary of expressed and biased orthologous and homoeologous genes in all 35 tissues from allotetraploid cotton. The corresponding data are from the *G. hirsutum* sequencing project [22]. A_t and D_t indicate the A and D subgenome, respectively. Y-axes indicate number of expressed gene pairs (left side) and ratio of biased to total gene pairs (right side). About 20–40% percent of genes showed expression bias. Significant ($P < 0.05$) differences by Fisher's exact test are indicated by asterisks.

Please cite this article as: L. Fang, et al., Asymmetric evolution and domestication in allotetraploid cotton (*Gossypium hirsutum* L.), The Crop Journal (2016), http://dx.doi.org/10.1016/j.cj.2016.07.001

with the pattern of gene expression reported in hexaploid bread wheat (*Triticum aestivum*) [20] and oilseed rape (*B. napus*) [21], but is in contrast to that in paleopolyploid maize [32], allotetraploid *Arabidopsis* [19] and mesopolyploid *B. rapa* [33]. However, 20–40% of homoeologous gene pairs show biased A- or D-subgenome expression depending on the tissue, owing to silencing or lower expression levels of one homoeolog [39–41]. The number of expressed genes biased toward the D subgenome was slightly greater than that of those biased toward the A subgenome consistently through most tissues (Fig. 2). In contrast, more transcription factor (TF) genes show expression bias toward the A homoeologs at 20 days post-anthesis (DPA) during fiber development (Fig. S1), such as MYB and MYB-related type TFs, indicating that the homoeolog expression bias varies among tissues depending on its biological function [22]. Given the ubiquity of homoeolog expression bias, it is likely to have biological impacts. The alterations in homoeolog expression bias accompany the merger of two diverged diploid genomes, owing to the combination of regulatory (*cis* or *trans*) and epigenetic interactions with the transcriptome network [42,43]. Moreover, the homoeolog expression bias is also sensitive to environmental conditions, meaning that certain conditions will alter or favor one of the homoeologs. For instance, abiotic stress can alter the homoeolog gene expression ratios in *G. hirsutum* [44]. Expression-level alterations set in motion by polyploidy are closely related to genome-wide subfunctionalization and neofunctionalization as a long-term response to polyploidy [45–47]. In addition to the homoeolog expression bias in cotton and other allotetraploids, there are further unbiased gene pairs, which contradict the individual gene expression bias. It is important to understand the possible underlying mechanisms of these biased and unbiased gene pairs.

6. Asymmetric domestication for cotton fiber biology and cotton adaptation

To gain insight into the genomic mysteries of cotton selection and domestication, a total of 811 positively selected genes (PSGs) were identified in allopolyploid cotton. These included 470 PSGs in the A subgenome, significantly more than the 341 PSGs identified in the D subgenome ($P < 0.01$, Fisher's exact test) [22]. Among these, 343 (72.9%) A subgenome PSGs and 248 (72.7%) D subgenome PSGs were expressed during fiber development (5–25 DPA), suggesting natural and human selection of traits leading to fiber improvement in cultivated cotton. The A-subgenome PSGs were significantly enriched in beta-D-glucan biosynthetic process (GO: 0006075), regulation of signal transduction (GO: 0009966), carbohydrate biosynthetic process (GO: 0016051) and sucrose biosynthetic process (GO: 0005986), which are known to affect fiber development. Interestingly, the D subgenome PSGs are enriched in response to superoxide (GO: 000303) and other abiotic stresses and were upregulated in response to four different stress treatments in Upland cotton. Moreover, 124 PSGs were upregulated in domesticated cotton in two key developmental stages (10 and 20 DPA), primary cell wall and secondary cell wall growth, in comparison with their expression in wild *G. hirsutum* races (*palmeri* and *yucatanense*), which have short fibers. Similarly, 70 A subgenome PSGs were associated mainly with fiber development and 54 D subgenome PSGs were associated with salt or osmotic stress tolerance and response to superoxide stress. These results show that fiber development-associated genes originating from the fiber-producing A-genome progenitor have played an important role in domestication. In contrast, the D subgenome contributes a wide range of adaptation traits to allotetraploid cotton.

7. Conclusion and future perspectives

One of the most important events in genomics is WGD, or polyploidy. More than 300,000 species of flowering plants have experienced variation in the size and complement of their genomic elements [48]. Genome sequences of allotetraploid cotton provide insights into the genomic response to polyploidy, including DNA-level and expression-level responses (Fig. 1). Genome-wide DNA-level alterations include asymmetric structural rearrangements, lost and disrupted genes, changes in evolution rate, and centromeric retrotransposon transfer; showing asymmetric evolution during allotetraploid cotton formation. Although there was no genome-wide expression bias in the expression level response in the A and D subgenomes, we observed homoeolog expression bias in different tissues, representing the underexplored mysteries of the transcriptomic dynamics of allopolyploid cotton. In addition to asymmetric evolution of genome structure and genes, whole-genome PSG analysis has shown that the A and D subgenomes contribute fiber and adaptation traits, respectively, thus deepening our understanding of why domesticated cotton produces longer, stronger, and finer fibers and shows wider adaptation to diverse environments. With this genomic information, we can further elucidate the mechanisms causing genomic alterations in genome structure and gene expression in natural polyploidy genomes, laying a solid foundation for genome-wide designer breeding and development of high-yielding, widely adapted plants with superior quality.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.cj.2016.07.001.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 3133058, 31290213), the Fundamental Research Funds for the Central Universities, the Priority Academic Program Development of Jiangsu Higher Education Institutions, and the 111 Project (B08025).

REFERENCES

[1] Z.J. Chen, Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids, Annu. Rev. Plant Biol. 58 (2007) 377–406.
[2] P.S. Soltis, D.E. Soltis, The role of hybridization in plant speciation, Annu. Rev. Plant Biol. 60 (2009) 561–588.

Please cite this article as: L. Fang, et al., Asymmetric evolution and domestication in allotetraploid cotton (*Gossypium hirsutum* L.), The Crop Journal (2016), http://dx.doi.org/10.1016/j.cj.2016.07.001

[3] A.R. Leitch, I.J. Leitch, Genomic plasticity and the diversity of polyploid plants, Science 320 (2008) 481–483.
[4] S.P. Otto, The evolutionary consequences of polyploidy, Cell 131 (2007) 452–462.
[5] Z.J. Chen, B.E. Scheffler, E. Dennis, B.A. Triplett, T.Z. Zhang, W.Z. Guo, X.Y. Chen, D.M. Stelly, P.D. Rabinowicz, C.D. Town, T. Arioli, C. Brubaker, R.G. Cantrell, J.M. Lacape, M. Ulloa, P. Chee, A.R. Gingle, C.H. Haigler, R. Percy, S. Saha, T. Wilkins, R.J. Wright, A. Van Deynze, Y.X. Zhu, S.X. Yu, I. Abdurakhmonov, I. Katageri, P.A. Kumar, M.u. Rahman, Y. Zafar, J.Z. Yu, R.J. Kohel, J.F. Wendel, A.H. Paterson, Toward sequencing cotton (*Gossypium*) genomes, Plant Physiol. 145 (2014) 1303–1310.
[6] A.E. Percival, J.F. Wendel, J.M. Stewart, Taxonomy and germplasm resources, in: C.W. Smith, J.T. Cothren (Eds.), Cotton: Origin, History, Technology, and Production, John Wiley & Sons, New York, USA 1999, pp. 33–63.
[7] J.F. Wendel, New World tetraploid cottons contain Old World cytoplasm, Proc. Natl. Acad. Sci. U. S. A. 86 (1989) 4132–4136.
[8] B. Hendrix, J.M. Stewart, Estimation of the nuclear DNA content of *Gossypium* species, Ann. Bot. 95 (2005) 789–797.
[9] F.G. Li, G.Y. Fan, K.B. Wang, F.M. Sun, Y.L. Yuan, G.L. Song, Q. Li, Z.Y. Ma, C.R. Lu, C.S. Zou, W.B. Chen, X.M. Liang, H.H. Shang, W.Q. Liu, C.C. Shi, G.H. Xiao, C.Y. Gou, W.W. Ye, X. Xu, X.Y. Zhang, H.L. Wei, Z.F. Li, G.Y. Zhang, J.Y. Wang, K. Liu, R.J. Kohel, R.G. Percy, J.Z. Yu, Y.X. Zhu, J. Wang, S.X. Yu, Genome sequence of the cultivated cotton *Gossypium arboreum*, Nat. Genet. 46 (2014) 567–572.
[10] A.H. Paterson, J.F. Wendel, H. Gundlach, H. Guo, J. Jenkins, D.C. Jin, D. Llewellyn, K.C. Showmaker, S.Q. Shu, J. Udall, M.J. Yoo, R. Byers, W. Chen, A. Doron-Faigenboim, M.V. Duke, L. Gong, J. Grimwood, C. Grover, K. Grupp, G.J. Hu, T.H. Lee, J.P. Li, L.F. Lin, T. Liu, B.S. Marler, J.T. Page, A.W. Roberts, E. Romanel, W.S. Sanders, E. Szadkowski, X. Tan, H.B. Tang, C.M. Xu, J.P. Wang, Z.N. Wang, D. Zhang, L. Zhang, H. Ashrafi, F. Bedon, J.E. Bowers, C.L. Brubaker, P.W. Chee, S. Das, A.R. Gingle, C.H. Haigler, D. Harker, L.V. Hoffmann, R. Hovav, D.C. Jones, C. Lemke, S. Mansoor, M.u. Rahman, L.N. Rainville, A. Rambani, U.K. Reddy, J.K. Rong, Y. Saranga, B.E. Scheffler, J.A. Scheffler, D.M. Stelly, B.A. Triplett, A. Van Deynze, M.F.S. Vaslin, V.N. Waghmare, S.A. Walford, R.J. Wright, E.A. Zaki, T.Z. Zhang, E.S. Dennis, K.F.X. Mayer, D.G. Peterson, D.S. Rokhsar, X.Y. Wang, J. Schmutz, Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres, Nature 492 (2012) 423–427.
[11] The Arabidopsis Genome Initiative, Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*, Nature 408 (2000) 796–815.
[12] T.P. Michael, S. Jackson, The first 50 plant genomes, Plant Genome 6 (2013) 547–562.
[13] P.S. Schnable, D. Ware, R.S. Fulton, J.C. Stein, F.S. Wei, S. Pasternak, C.Z. Liang, J.W. Zhang, L. Fulton, T.A. Graves, P. Minx, A.D. Reily, L. Courtney, S.S. Kruchowski, C. Tomlinson, C. Strong, K. Delehaunty, C. Fronick, B. Courtney, S.M. Rock, E. Belter, F.Y. Du, K. Kim, R.M. Abbott, M. Cotton, A. Levy, P. Marchetto, K. Ochoa, S.M. Jackson, B. Gillam, W.Z. Chen, L. Yan, J. Higginbotham, M. Cardenas, J. Waligorski, E. Applebaum, L. Phelps, J. Falcone, K. Kanchi, T. Thane, A. Scimone, N. Thane, J. Henke, T. Wang, J. Ruppert, N. Shah, K. Rotter, J. Hodges, E. Ingenthron, M. Cordes, S. Kohlberg, J. Sgro, B. Delgado, K. Mead, A. Chinwalla, S. Leonard, K. Crouse, K. Collura, D. Kudrna, J. Currie, R.F. He, A. Angelova, S. Rajasekar, T. Mueller, R. Lomeli, G. Scara, A. Ko, K. Delaney, M. Wissotski, G. Lopez, D. Campos, M. Braidotti, E. Ashley, W. Golser, H. Kim, S. Lee, J.K. Lin, Z. Dujmic, W. Kim, J. Talag, A. Zuccolo, C.Z. Fan, A. Sebastian, M. Kramer, L. Spiegel, L. Nascimento, T. Zutavern, B. Miller, C. Ambroise, S. Muller, W. Spooner, A. Narechania, L. Ren, S. Wei, S. Kumari, B. Faga, M.J. Levy, L. McMahan, P. Van Buren, M.W. Vaughn, K. Ying, C.T. Yeh, S.J. Emrich, Y. Jia, A. Kalyanaraman, A.P. Hsia, W.B. Barbazuk, R.S. Baucom, T.P. Brutnell, N.C. Carpita, C. Chaparro, J.M. Chia, J.M. Deragon, J.C. Estill, Y. Fu, J.A. Jeddeloh, Y.J. Han, H. Lee, P.H. Li, D.R. Lisch, S.Z. Liu, Z.J. Liu, D.H. Nagel, M.C. McCann, P. SanMiguel, A.M. Myers, D. Nettleton, J. Nguyen, B.W. Penning, L. Ponnala, K.L. Schneider, D.C. Schwartz, A. Sharma, C. Soderlund, N.M. Springer, Q. Sun, H. Wang, M. Waterman, R. Westerman, T.K. Wolfgruber, L.X. Yang, Y. Yu, L.F. Zhang, S.G. Zhou, Q.H. Zhu, J.L. Bennetzen, R.K. Dawe, J.M. Jiang, N. Jiang, G.G. Presting, S.R. Wessler, S. Aluru, R.A. Martienssen, S.W. Clifton, W.R. McCombie, R.A. Wing, R.K. Wilson, The B73 maize genome: complexity, diversity, and dynamics, Science 326 (2009) 1112–1115.
[14] J. Schmutz, S.B. Cannon, J. Schlueter, J.X. Ma, T. Mitros, W. Nelson, D.L. Hyten, Q.J. Song, J.J. Thelen, J.L. Cheng, D. Xu, U. Hellsten, G.D. May, Y. Yu, T. Sakurai, T. Umezawa, M.K. Bhattacharyya, D. Sandhu, B. Valliyodan, E. Lindquist, M. Peto, D. Grant, S.Q. Shu, D. Goodstein, K. Barry, M. Futrell-Griggs, B. Abernathy, J.C. Du, Z.X. Tian, L.C. Zhu, N. Gill, T. Joshi, M. Libault, A. Sethuraman, X.C. Zhang, K. Shinozaki, H.T. Nguyen, R.A. Wing, P. Cregan, J. Specht, J. Grimwood, D. Rokhsar, G. tacey, R.C. Shoemaker, S.A. Jackson, Genome sequence of the palaeopolyploid soybean, Nature 463 (2010) 178–183.
[15] R. Brenchley, M. Spannagl, M. Pfeifer, G.L.A. Barker, R. D'Amore, A.M. Allen, N. McKenzie, M. Kramer, A. Kerhornou, D. Bolser, S. Kay, D. Waite, M. Trick, I. Bancroft, Y. Gu, N.X. Huo, M.C. Luo, S. Sehgal, S. Kianian, B. Gill, O. Anderson, P. Kersey, J. Dvorak, R. McCombie, A. Hall, K.F.X. Mayer, K.J. Edwards, M.W. Bevan, N. Hall, Analysis of the bread wheat genome using whole-genome shotgun sequencing, Nature 491 (2012) 705–710.
[16] M.R. Woodhouse, J.C. Schnable, B.S. Pedersen, E. Lyons, D. Lisch, S. Subramaniam, M. Freeling, Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homeologs, PLoS Biol. 8 (2010), e1000409.
[17] M. Feldman, L. Bao, G. Segal, S. Abbo, A.A. Levy, J.M. Vega, Rapid elimination of low-copy DNA sequences in polyploid wheat: a possible mechanism for differentiation of homoeologous chromosomes, Genetics 147 (1997) 1381–1387.
[18] R.T. Gaeta, J.C. Pires, F. Iniguezuy, E. Leon, T.C. Osborn, Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype, Plant Cell 19 (2007) 3403–3417.
[19] J.L. Wang, L. Tian, H.S. Lee, N.E. Wei, H.M. Jiang, B. Watson, A. Madlung, T.C. Osborn, R.W. Doerge, L. Comai, Z.J. Chen, Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids, Genetics 172 (2006) 507–517.
[20] International Wheat Genome Sequencing Consortium, A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome, Science 345 (2014) 6194.
[21] B. Chalhoub, F. Denoeud, S.Y. Liu, I.A.P. Parkin, H.B. Tang, X.Y. Wang, J. Chiquet, H. Belcram, C.B. Tong, B. Samans, M. Corréa, C. Da Silva, J. Just, C. Falentin, C.S. Koh, I. Le Clainche, M. Bernard, P. Bento, B. Noel, K. Labadie, A. Alberti, M. Charles, D. Arnaud, H. Guo, C. Daviaud, S. Alamery, K. Jabbari, M.X. Zhao, P.P. Edger, H. Chelaifa, D. Tack, G. Lassalle, I. Mestiri, N. Schnel, M.C. Le Paslier, G.Y. Fan, V. Renault, P.E. Bayer, A.A. Golicz, S. Manoli, T.H. Lee, V.H.D. Thi, S. Chalabi, Q. Hu, C.C. Fan, R. Tollenaere, Y.H. Lu, C. Battail, J.X. Shen, C.H.D. Sidebottom, X.F. Wang, A. Canaguier, A. Chauveau, A. Bérard, G. Deniot, M. Guan, Z.S. Liu, F.M. Sun, Y.P. Lim, E. Lyons, C.D. Town, I. Bancroft, X.W. Wang, J.L. Meng, J.X. Ma, J.C. Pires, G.J. King, D. Brunel, R. Delourme, M. Renard, J.M. Aury, K.L. Adams, J. Batley, R.J. Snowdon, J. Tost, D. Edwards, Y.M. Zhou, W. Hua, A.G. Sharpe, A.H. Paterson, C.Y. Guan, P. Wincker, Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome, Science 345 (2014) 950.

Please cite this article as: L. Fang, et al., Asymmetric evolution and domestication in allotetraploid cotton (*Gossypium hirsutum* L.), The Crop Journal (2016), http://dx.doi.org/10.1016/j.cj.2016.07.001

[22] T.Z. Zhang, Y. Hu, W.K. Jiang, L. Fang, X.Y. Guan, J.D. Chen, J.B. Zhang, C.A. Saski, B.E. Scheffler, D.M. Stelly, A.M. Hulse-Kemp, Q. Wan, B.L. Liu, C.X. Liu, S. Wang, M.Q. Pan, Y.K. Wang, D.W. Wang, W.X. Ye, L.J. Chang, W.P. Zhang, Q.X. Song, R.C. Kirkbride, X.Y. Chen, E. Dennis, D.J. Llewellyn, D.G. Peterson, P. Thaxton, D.C. Jones, Q. Wang, X.Y. Xu, H. Zhang, H.T. Wu, L. Zhou, G.F. Mei, S.Q. Chen, Y. Tian, D. Xiang, X.H. Li, J. Ding, Q.Y. Zuo, L.N. Tao, Y.C. Liu, J. Li, Y. Lin, Y.Y. Hui, Z.S. Cao, C.P. Cai, X.F. Zhu, Z. Jiang, B.L. Zhou, W.Z. Guo, R.Q. Li, Z.J. Chen, Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fibre improvement, Nat. Biotechnol. 33 (2015) 531–537.

[23] F.G. Li, G.Y. Fan, C.R. Lu, G.H. Xiao, C.S. Zou, R.J. Kohel, Z.Y. Ma, H.H. Shang, X.F. Ma, J.Y. Wu, X.M. Liang, G. Huang, R.G. Percy, K. Liu, W.H. Yang, W.B. Chen, X.M. Du, C.C. Shi, Y.L. Yuan, W.W. Ye, X. Liu, X.Y. Zhang, W.Q. Liu, H.L. Wei, S.J. Wei, G.D. Huang, X.L. Zhang, S.J. Zhu, H. Zhang, F.M. Sun, X.F. Wang, J. Liang, J.H. Wang, Q. He, L.H. Huang, J. Wang, J.J. Cui, G.L. song, K.B. Wang, X. Xu, J.Z. Yu, Y.X. Zhu, S.X. Yu, Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution, Nat. Biotechnol. 33 (2015) 524–530.

[24] Z.J. Chen, Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids, Annu. Rev. Plant Biol. 58 (2007) 377–406.

[25] K. Adams, J.F. Wendel, Polyploidy and genome evolution in plants, Curr. Opin. Plant Biol. 8 (2005) 135–141.

[26] S. Wang, J.D. Chen, W.P. Zhang, Y. Hu, L.J. Chang, L. Fang, Q. Wang, F.N. Lv, H.T. Wu, Z.F. Si, S.Q. Chen, C.P. Cai, X.F. Zhu, B.L. Zhou, W.Z. Guo, T.Z. Zhang, Sequence-based ultra-dense genetic and physical maps reveal structural variations of allopolyploidy cotton genomes, Genome Biol. 16 (2015) 108.

[27] I.J. Leitch, A.R. Leitch, Genome Size Diversity and Evolution in Land Plants, in: I.J. Leitch, J. Greilhuber, J. Doležel, J.F. Wendel (Eds.), Plant Genome Diversity Volume 2: Physical Structure, Behaviour and Evolution of Plant Genomes, Springer, Vienna, Austria 2013, pp. 307–322.

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

[29] S.Y. Liu, Y.M. Liu, X.H. Yang, C.B. Tong, D. Edwards, I.A.P. Parkin, M.X. Zhao, J.X. Ma, J.Y. Yu, S.M. Huang, X.Y. Wang, J.Y. Wang, K. Lu, Z.Y. Fang, I. Bancroft, T.J. Yang, Q. Hu, X.F. Wang, Z. Yue, H.J. Li, L.F. Yang, J. Wu, Q. Zhou, W.X. Wang, G.J. King, J.C. Pires, C.X. Lu, Z.Y. Wu, P. Sampath, Z. Wang, H. Guo, S.K. Pan, L.M. Yang, J.M. Min, D. Zhang, D.C. Jin, W.S. Li, H. Belcram, J.X. Tu, M. Guan, C.K. Qi, D.Z. Du, J.N. Li, L.C. Jiang, J. Batley, A.G. Sharpe, B.S. Park, P. Ruperao, F. Cheng, N.E. Waminal, Y. Huang, C.H. Dong, L. Wang, J.P. Li, Z.Y. Hu, M. Zhuang, Y. Huang, J.Y. Huang, J.Q. Shi, D.S. Mei, J. Liu, T.H. Lee, J.P. Wang, H.Z. Jin, Z.Y. Li, X. Li, J.F. Zhang, L. Xiao, Y.M. Zhou, Z.S. Liu, X.Q. Liu, R. Qin, X. Tang, W.B. Liu, Y.P. Wang, Y.Y. Zhang, J. Lee, H.H. Kim, F. Denoeud, X. Xu, X.M. Liang, W. Hua, X.W. Wang, J. Wang, B. Chalhoub, A.H. Paterson, The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes, Nat. Commun. 5 (2014) 3930.

[30] E. Casacuberta, J. González, The impact of transposable elements in environmental adaptation, Mol. Ecol. 22 (2013) 1503–1517.

[31] K.B. Wang, Z.W. Wang, F.G. Li, W.W. Ye, J.Y. Wang, G.L. Song, Z. Yue, L. Cong, H.H. Shang, S.L. Zhu, C.S. Zou, Q. Li, Y.L. Yuan, C.R. Lu, H.L. Wei, C.Y. Gou, Z.Q. Zheng, Y. Yin, X.Y. Zhang, K. Liu, B. Wang, C. Song, N. Shi, R.J. Kohel, R.G. Percy, J.Z. Yu, Y.X. Zhu, J. Wang, S.X. Yu, The draft genome of a diploid cotton *Gossypium raimondii*, Nat. Get. 44 (2012) 1098–1103.

[32] J.C. Schnable, N.M. Springer, M. Feeling, Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 4069–4074.

[33] F. Cheng, J. Wu, L. Fang, S. Sun, B. Liu, et al., Biased gene fractionation and dominant gene expression among the subgenomes of *Brassica rapa*, PLoS One 7 (2012).

[34] S. Renny-Byfield, L. Gong, J.P. Gallagher, J.F. Wendel, Persistence of subgenomes in paleopolyploid cotton after 60 My of evolution, Mol. Biol. Evol. 32 (2015) 1063–1071.

[35] S. Wang, J. Wang, J. Jiang, Q. Zhang, Mapping of centromeric regions on the molecular linkage map of rice (*Oryza sativa* L.) using centromere-associated sequences, Mol. Gen. Genet. 263 (2000) 165–172.

[36] S. Luo, J. Mach, B. Abramson, R. Ramire, R. Schurr, P. Barone, G. Copenhaver, O. Folkerts, The cotton centromere contains a Ty3-gypsy-like LTR retroelement, PLoS One 7 (2012), e35261.

[37] W.P. Zhang, Y.J. Cao, K. Wang, T. Zhang, J.D. Chen, M.Q. Pan, Q. Wang, S.L. Feng, W.Z. Guo, B.L. Zhou, T.Z. Zhang, Identification of centromeric regions on the linkage map of cotton using centromere-related repeats, Genomics 104 (2014) 587–593.

[38] K.L. Adams, L. Flagel, J.F. Wendel, Responses of the cotton genome to polyploidy, in: A. Paterson (Ed.), Genetics and Genomics of Cotton, Springer, New York 2009, pp. 419–429.

[39] K. Adams, R. Cronn, R. Percifield, J.F. Wendel, Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 4649–4654.

[40] L.E. Flagel, J.F. Wendel, Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation, New Phytol. 186 (2010) 184–193.

[41] M. Yoo, J. Wendel, Comparetive evolutionary and development dynamics of the cotton (*Gossypium hirsutum*) fiber transcriptome, PLoS Genet. 10 (2014), e1004073.

[42] R.J.A. Buggs, L.J. Zhang, N. Miles, J.A. Tate, L. Gao, W. Wei, P.S. Schnable, W.B. Barbazuk, P.S. Soltis, D.E. Soltis, Transcriptomic shock generates evolutionary novelty in a newly formed, natural allopolyploid plant, Curr. Biol. 21 (2011) 551–556.

[43] M. Yoo, E. Szadkowski, J.F. Wendel, Homoeolog expression bias and expression level dominance in allopolyploid cotton, Heredity 110 (2013) 171–180.

[44] S. Dong, K.L. Adams, Differential contributions to the transcriptome of duplicated genes in response to abiotic stresses in natural and synthetic polyploids, New Phytol. 190 (2011) 1045–1057.

[45] S. Liu, K.L. Adams, Dramatic change in function and expression pattern of a gene duplicated by polyploidy created a paternal effect gene in the *Brassicaceae*, Mol. Biol. Evol. 27 (2010) 2817–2828.

[46] T.E. Hughes, J.A. Langdale, S. Kelly, The impact of widespread regulatory neofunctionalization on homeolog gene evolution following whole-genome duplication in maize, Genome Res. 24 (2014) 1348–1355.

[47] S. Renny-Byfield, J.P. Gallagher, C.E. Grover, E. Szadkowski, J.T. Page, J.A. Udall, X.Y. Wang, A.H. Paterson, J.F. Wendel, Ancient gene duplicates in *Gossypium* (cotton) exhibit near-complete expression divergence, Genome Biol. Evol. 6 (2014) 559–571.

[48] J.F. Wendel, S. Jackson, B. Meyers, R. Wing, Evolution of plant genome architecture, Genome Biol. 17 (2016) 37.

Please cite this article as: L. Fang, et al., Asymmetric evolution and domestication in allotetraploid cotton (*Gossypium hirsutum* L.), The Crop Journal (2016), http://dx.doi.org/10.1016/j.cj.2016.07.001