New Phytologist

Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation

Lex E. Flagel and Jonathan F. Wendel

Department of Ecology, Evolution and Organismal Biology, Iowa State University, Ames, IA 50011, USA

Author for correspondence:
Jonathan F. Wendel
Tel: +1 515 294 7172
Email jfw@iastate.edu

Received: *30 August 2009*
Accepted: *11 October 2009*

New Phytologist (2010) **186**: 184–193
doi: 10.1111/j.1469-8137.2009.03107.x

Key words: gene expression, *Gossypium* (cotton), homoeolog, microarray, polyploidy.

Summary

• Here, we describe the evolution of gene expression among a diversified cohort of five allopolyploid species in the cotton genus (*Gossypium*). Using this phylogenetic framework and comparisons with expression changes accompanying F_1 hybridization, we provide a temporal perspective on expression diversification following a shared genome duplication.

• Global patterns of gene expression were studied by the hybridization of petal RNAs to a custom microarray. This platform measures total expression for *c.* 42 000 duplicated genes, and genome-specific expression for *c.* 1400 homoeologs (genes duplicated by polyploidy).

• We report homoeolog expression bias favoring the allopolyploid D genome over the A genome in all species (among five polyploid species, D biases ranging from *c.* 54 to 60%), in addition to conservation of biases among genes. Furthermore, we find surprising levels of transgressive up- and down-regulation in the allopolyploids, a diminution of the level of bias in genomic expression dominance but not in its magnitude, and high levels of rate variation among allotetraploid species.

• We illustrate how phylogenetic and temporal components of expression evolution may be partitioned and revealed following allopolyploidy. Overall patterns of expression evolution are similar among the *Gossypium* allotetraploids, notwithstanding a high level of interspecific rate variation, but differ strikingly from the direction of genomic expression dominance patterns in the synthetic F_1 hybrid.

Introduction

As testified by the many papers in this volume, polyploidy, or whole genome duplication, is a prevalent feature among angiosperm species (Wendel, 2000; Comai, 2005; Leitch & Leitch, 2008). Emerging genomic data have shed light on the ancient and recurrent history of polyploidy among the angiosperms (Barker *et al.*, 2008; Tang *et al.*, 2008; Soltis *et al.*, 2009). Because polyploidy involves the duplication of the entire genome, its effect on genomic organization can be extensive (Comai, 2005), including well-documented cases of structural and epigenetic modifications (Shaked *et al.*, 2001; Gaeta *et al.*, 2007; Buggs *et al.*, 2009; Ni *et al.*, 2009; Tate *et al.*, 2009), as well as changes in gene expression patterns (Bottley *et al.*, 2006; Hegarty *et al.*, 2006; Wang *et al.*, 2006; Flagel *et al.*, 2008; Hovav *et al.*, 2008; Rapp *et al.*, 2009). Furthermore, some of these genome-wide changes have been linked to phenotypic variation (Pires *et al.*, 2004; Gaeta *et al.*, 2007; Ni *et al.*, 2009), providing direct support for the long-held notion that polyploidy can be an important driver of phenotypic evolution.

The establishment of a new allopolyploid species is not a trivial feat. First, all allopolyploids face several immediate genomic challenges, including the merger of divergent genomes, the resolution of potentially conflicting developmental signals and new or possibly accidental interactions with organellar genomes, in addition to overcoming the reproductive barriers associated with polyploidy (Wendel, 2000; Comai, 2005). Following this, and owing to their redundant genomic architecture, allopolyploid genomes then face several interesting and potentially dramatic evolutionary resolutions. These include the genomic decay of

duplicate genes either in the form of genomic fragment loss (Shaked *et al.*, 2001; Tate *et al.*, 2009) or mutational obliteration (pseudogenization), genomic partitioning of ancestral functions (subfunctionalization; Force *et al.*, 1999) or the possibility of a chance beneficial mutation conferring new functionality (neofunctionalization; Ohno, 1970). These outcomes are not mutually exclusive (Conant & Wolfe, 2008), and most probably require evolutionary time-scales, and can be distorted by additional genomic disruptions, such as further hybridization and/or polyploidization leading to the accumulation of additional genomic content, yielding higher ploidies and additional genomic complexity [e.g. *Spartina anglica*, sugarcane (*Saccharum officinarum*) or wheat (*Triticum aestivum*)]. In the absence of hybridization or additional rounds of polyploidization, nascent polyploids can undergo divergence and spawn cladogenesis, as has happened in hundreds of genera throughout the angiosperms. As this special edition of *New Phytologist* demonstrates, the polyploid research community has made major inroads into the study of the genomic consequences of polyploidy. Despite this progress, many important questions remain. The study presented here addresses one of these questions using a model system from the cotton genus, namely, how is gene expression among newly co-resident genomes affected during the lengthy process of allopolyploid diversification?

The organismal context for this analysis is as follows: 1–2 million years ago, allopolyploidization within the genus *Gossypium* resulted in a new allotetraploid lineage containing diploid genomes from both the Old World A genome and New World D genome (Senchina *et al.*, 2003; Wendel & Cronn, 2003). Since that time, species containing this favorable genomic combination have spread throughout the tropical and subtropical portions of the New World and have diversified into five extant allotetraploid species (Wendel & Cronn, 2003), although a sixth species, *G. ekmanianum*, has been proposed recently (Krapovickas & Seijo, 2008). The presence of shared allopolyploid-specific nucleotide polymorphisms within these species indicates that they probably evolved from a single polyploidy event and, as a consequence, have left a traceable phylogenetic history which has been revealed by previous studies (Wendel *et al.*, 1994; Small *et al.*, 1998) (Fig. 1a).

The evolutionary framework provided by the five natural *Gossypium* allotetraploids offers an excellent opportunity to study replicated evolutionary trajectories following the combination of diversified genomes. In addition to their compelling natural history, two allotetraploid cottons, *G. hirsutum* and *G. barbadense*, are primary contributors of natural fiber for use in the textile and apparel industries, making it agriculturally and economically important to understand their evolutionary history. The study of these allopolyploids has benefited from considerable genomic resources, including a sizable expressed sequence tag (EST)

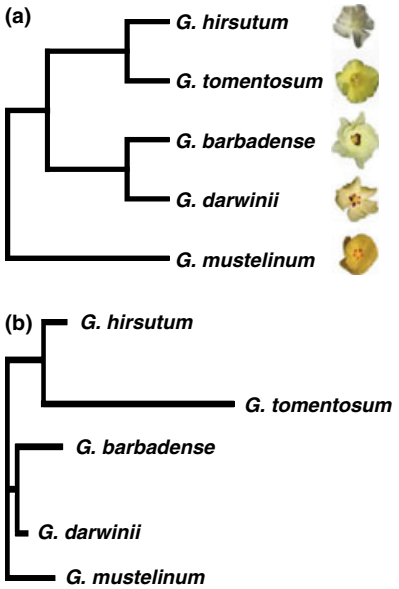

Fig. 1 *Gossypium* allotetraploid phylogeny and 'expression phylogram'. (a) The phylogeny of the five *Gossypium* allotetraploids, including an image of their flowers at maturity. (b) A phylogeny of the same species, where the branch length represents the extent of homoeologous expression divergence among 1383 genes.

collection (Udall *et al.*, 2006a), with ESTs from both model diploid parents (A genome: *G. arboreum*; D genome: *G. raimondii*), which are not the exact progenitors of the natural allotetraploid cottons, but are the closest modern representatives, as well as the allotetraploid *G. hirsutum* (Table 1). This genomic resource has been used to create a novel microarray platform, which can be used to explore global gene expression levels among *c.* 42 000 genes using probes targeted at conserved genic regions of the A and D cotton genomes, and homoeologous (genes duplicated by polyploidy) expression levels for *c.* 1400 genes using pairs of probes differentiated by a genome-specific single nucleotide polymorphism (Udall *et al.*, 2006b; Flagel *et al.*, 2008).

Using this microarray platform, several key findings have been made regarding polyploidy in *Gossypium*. Most relevant to the present study, we have shown previously that both genomic merger and allopolyploid evolution play important roles in homoeolog expression evolution (Flagel *et al.*, 2008), and that homoeolog expression is biased in favor of the D genome in *G. hirsutum* in both petal and fiber tissues (Flagel *et al.*, 2008; Hovav *et al.*, 2008). Following these initial findings regarding homoeologous expression, continued work with this microarray platform has highlighted a form of genomic expression dominance, whereby the allotetraploid assumes an expression state of the D genome parent significantly more often than it does the A genome parent, regardless of whether this state is up- or down-regulation (Rapp *et al.*, 2009). Beyond these studies in *Gossypium*, work in allopolyploid wheat (Bottley

New Phytologist

Table 1 *Gossypium* taxa used in this study

Species' name	Genome designation	Accession	Ploidy level	Geographic origin of species	Petal harvest dates[1]
G. arboreum	A_2	cv. AKA-8410	Diploid	Africa	May 2–June 5
G. raimondii	D_5	Accession unnamed	Diploid	South America	Mar 9–Apr 6
G. hirsutum	AD_1	cv. Maxxa	Allotetraploid	Mexico/Central America	May 9–May 29
G. barbadense	AD_2	cv. Pima S7	Allotetraploid	South America	May 8–May 31
G. tomentosum	AD_3	WT936	Allotetraploid	Hawaii	Apr 4–Apr 19
G. mustelinum	AD_4	15C	Allotetraploid	NE Brazil	Jan 24–Feb 24
G. darwinii	AD_5	PW45	Allotetraploid	Galapagos Islands	Jan 24–Feb 11
F_1 hybrid	$A_2♀ \times D_5♂$	Accession unnamed	Diploid	Synthetic hybrid	Jan 25–Mar 3

[1], All harvest dates are from the year 2006.

et al., 2006; Bottley & Koebner, 2008; Pumphrey *et al.*, 2009) and *Tragopogon* (Tate *et al.*, 2006) has further demonstrated a considerable frequency of biases in the genomic contribution among homoeologs, and work in hybrids between *Arabidopsis* autotetraploids has shown global down-regulation of the *A. thaliana* genome in favor of the *A. arenosa* genome (Wang *et al.*, 2006), which could be considered as another form of genomic dominance. Together, these observations are beginning to confirm the notion that the genomic disruptions associated with allopolyploidy may contribute considerably to gene expression evolution within established and nascent polyploids (Osborn *et al.*, 2003; Chen, 2007; Paun *et al.*, 2007; Doyle *et al.*, 2008).

Here, we extend the scope of earlier findings by demonstrating significant levels of expression evolution among a diversified collection of natural allopolyploid species, further refining our temporal perspective on expression evolution and revealing extraordinary variation in the rate of expression evolution among a diversifying lineage. We also show aspects of expression evolution that are shared among the five natural allotetraploid cotton species and that are different from those exhibited in recently formed synthetic intergenomic hybrids.

Materials and Methods

Plant materials, RNA extraction and microarray preparation

Our study utilized five natural *Gossypium* allotetraploids, as well as their model A and D genome diploid progenitors and a diploid F_1 hybrid made by crossing the diploid progenitors (Table 1). A synthetic allopolyploid deriving from the model A and D genome diploid progenitors would add an additional dimension to our study. However, despite considerable effort, we have been unable to generate this accession. Replicates of all *Gossypium* plant materials were grown under controlled glasshouse conditions in the Pohl Conservatory at Iowa State University, USA. All plants were grown in a randomized block design with three biological replicates under full sunlight supplemented with sodium lighting for 10 h d^{-1}. Petals were selected as a focal tissue because flower maturation and petal opening on the day of anthesis follow a highly canalized trajectory among the *Gossypium* species studied, thus giving us the best possible opportunity to synchronize tissues collected on different days and among different species. Petal tissues were harvested from these accessions between January and June of 2006 (dates are provided for each species in Table 1), between *c.* 10:00 h and 12:00 h, which corresponds to the time of full petal expansion for all species. All petal tissues were snap frozen in liquid nitrogen and stored at −80°C.

Multiple flowers (> 3) were pooled by plant from three plants to form three biological replicates, which were then subjected to RNA extraction following a modified hot borate procedure (Wan & Wilkins, 1994). All RNA extractions were performed by replicate once all petals had been collected. Following extraction, RNA samples were run on a Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) to assess degradation. Finally, total RNA extracts were sent to Roche NimbleGen (Madison, WI, USA) for labeling and hybridization to a custom *Gossypium* microarray platform (microarray design details found in Flagel *et al.*, 2008). Briefly, this microarray features two classes of probes, including 7574 *c.* 35-mer pairs of A and D genome-specific probes (each containing a genome-specific single nucleotide polymorphism at their central base; targeting 1383 contigs), which have been demonstrated previously to possess diagnostic ability in assessing levels of A and D genome expression within an A by D genome F_1 hybrid (*G. arboreum* × *G. raimondii*) and allopolyploid *G. hirsutum* (Flagel *et al.*, 2008), as well as 297 206 *c.* 60-mer generic probes (conserved between the A and D genome; targeting 42 459 contigs), which have been utilized to detect global expression, without homoeolog specificity (Chaudhary *et al.*, 2008; Rapp *et al.*, 2009). Thus, this microarray platform makes it possible to measure total expression for *c.* 80% of the estimated genic content of the cotton genome (Rabinowicz *et al.*, 2005) and, for a smaller subset of genes, the platform can also detect the proportions of A and D genome contribution.

Statistical analysis

Microarray hybridizations were performed in two sets: the first set included *G. arboreum*, *G. ramondii*, *G. hirsutum*, the F_1 hybrid and an equimolar mix of RNA pools from the model diploid progenitors [*G. arboreum* (A genome) and *G. raimondii* (D genome)]. This first set was used to validate the utility of our microarray platform (Flagel *et al.*, 2008) and, following its success, the second set comprising the remaining four natural allotetraploid species (*G. barbadense*, *G. darwinii*, *G. mustelinum* and *G. tomentosum*; Table 1) was hybridized. These datasets were combined, using the conservative normalization procedures outlined in the next two paragraphs. All raw microarray data were extracted into two working files, one for the *c.* 35-mer genome-specific probes and one for the *c.* 60-mer generic probes. These genome-specific and generic datasets were normalized and subjected to statistical analysis separately, as they represent dissimilar probe types, each addressing different aspects of gene expression.

For the 7574 diagnostic genome-specific probe pairs (see Flagel *et al.*, 2008 for details regarding diagnostic probe selection), all raw values were natural log transformed and quantile normalized. Following this, the expression values of each pair were converted to the difference between the natural logs of the A and D genome probes [$\log_e(A_{probe}) - \log_e(D_{probe})$; hereafter referred to as the log ratio]. These log ratio values were reduced to the 1383 contigs they represent by calculating a robust average of all probe pairs for each contig using Tukey's biweight method. Finally, contig-level expression differences were determined using a linear model which included genotype and replication effects. This model was used to contrast the five natural allotetraploid species and the F_1 hybrid to the parental mix. The *P* values derived from this contrast were corrected for multiple testing using the method of Storey & Tibshirani (2003). Significance was assessed from the resulting *q* values using a false discovery rate threshold of $q \leq 0.15$. This threshold was arrived at by first estimating the number of true nulls using the method described by Nettleton *et al.* (2006), which is applied to the *P* value distribution to derive an estimate of the expected number of true null tests (no change in expression), a value that can be used to guide threshold selection when compared with estimates of statistically equivalent expression (here equivalent expression is operationally defined as the absence of statistically significant A or D genome biases) at various *q*-value thresholds (Table 2). Using this approach, we found that a *q*-value threshold of ≤ 0.15 was a good compromise between the expected number of true nulls and the observed cases of equivalent expression for all accessions. The results from the *q*-value thresholds, $q \leq 0.05$ and $q \leq 0.1$, can also be found in Table S1 (see Supporting Information).

The analysis of expression from the 297 206 *c.* 60-mer generic probes has been described previously by Rapp *et al.* (2009), and follows a general outline similar to that above.

Table 2 Categorization of A and D genome biases and equivalent contribution to the transcriptome for 1383 homoeologous/allelic gene pairs, including the estimate of true nulls ('Est. True H_0'; compare with the 'Equivalent' category), and the intersection of gene lists for: (1) all F_1 species including the hybrid; and (2) for only the five allotetraploid species, including in both cases their totals

Accession	A-biased	D-biased	Equivalent	Total	Est. true H_0
F_1	153	334	896	1383	905
G. hirsutum	455	570	358	1383	504
G. tomentosum	552	666	165	1383	352
G. barbadense	486	720	177	1383	391
G. darwinii	373	441	569	1383	591
G. mustelinum	292	370	721	1383	730
All species' intersection[1]	30**	79**	8*	117	NA
Only allotet. intersection[1]	176**	208**	12*	396	NA

[1], The observed extent of intersection among gene lists was tested relative to the level of intersection expected to occur by chance using a chi-squared test. All observed values were significantly greater than expected by chance (*, $P < 0.05$; **, $P < 0.001$).

The expression values were natural log transformed, quantile normalized and reduced to 42 459 contigs using Tukey's biweight method. Following this, expression differences were detected after fitting a linear model which included genotype and replication effects. *P* values from these contrasts were converted to *q* values using the method of Storey & Tibshirani (2003), and a threshold of $q \leq 0.05$ was used to assess significance to allow direct comparison with the results of Rapp *et al.* (2009).

Validation

We validated our microarray estimates of homeolog expression for 14 genes using a sensitive single nucleotide polymorphism-specific Sequenom (San Diego, CA, USA) mass spectrometry platform that was initially described for use in maize by Stupar & Springer (2006), and has a proven utility for estimating homoeologous expression ratios in *Gossypium* (Chaudhary *et al.*, 2009a) and for the validation of our custom *Gossypium* microarray (Flagel *et al.*, 2008; Hovav *et al.*, 2008). Using this platform, we compared homoeolog expression ratios between the microarray and mass spectrometry platforms for *G. barbadense*, *G. darwinii*, *G. mustelinum* and *G. tomentosum* (Fig. S1, see Supporting Information); the *G. hirsutum* and F_1 hybrid microarray expression estimates have been validated previously (Flagel *et al.*, 2008). The validations show significant correlations between the microarray and mass spectrometry estimates for *G. darwinii*, *G. mustelinum* and *G. tomentosum* (Pearson's $r = 0.525$, 0.535 and 0.54; $P = 0.053$, 0.048 and 0.046, respectively), and a nonsignificant, although moderate, correlation for *G. barbadense* (Pearson's $r = 0.366$,

$P = 0.19$). Despite the nonsignificant correlation for *G. barbadense*, these results confirm the quality of our microarray data, when we take into account the major technological differences between microarray and mass spectrometry platforms and a considerable history of validated results for this platform when applied to *Gossypium* (Chaudhary *et al.*, 2008, 2009a,b; Flagel *et al.*, 2008; Hovav *et al.*, 2008; Rapp *et al.*, 2009).

Microarray data deposition

Original microarray data files have been deposited in compliance with MIAME standards on the NCBI GEO website, and can be found under the dataset record GSE17927.

Results

Comparision of homoeolog expression biases between allotetraploid cottons

Using a well-established phylogeny and genomic history for five allopolyploid *Gossypium* species, we have assessed the ratio of homoeologous contributions to the transcriptome of petal tissues among 1383 duplicate gene pairs. After applying a false discovery rate threshold of 0.15 for significance testing, we tabulated the A-biased (significantly more A genome expression than the 1 : 1 parental mix), D-biased (significantly more D genome expression than the 1 : 1 parental mix) and equivalently expressed genes for each of the five allotetraploid *Gossypium* species and a synthetic F_1 hybrid (Table 2). The 1 : 1 parental RNA mix represents a best approximation of the anticipated expression state within the allotetraploids and F_1 in the absence of gene expression evolution. From our results, it is clear that all species show considerable deviations from this parental mix, with each species showing a substantial number of genes with both A and D genome biases. As was the case in our previous study (Flagel *et al.*, 2008), the F_1 hybrid shows fewer biases overall than do any of the five allotetraploids. In addition, among the allotetraploids, there is extraordinary variation in the number of departures from equivalence, with *G. mustelinum*, a wild species and the most basal of the *Gossypium* allotetraploids (Fig. 1a), showing the least divergence from the null expectation of expression equivalent to the parental mix, and *G. tomentosum* (a wild Hawaiian Island endemic) and *G. barbadense* (a domesticated South American species) showing the greatest levels of homoeolog expression bias (Table 2). Also consistent with our previous studies of petal and fiber tissues (Flagel *et al.*, 2008; Hovav *et al.*, 2008), all five allotetraploids and the F_1 hybrid show a greater number of paternal D-biased genes than maternal A-biased genes.

Because the *Gossypium* allotetraploids have a known phylogenetic history (Fig. 1a), it is possible to visualize homoeologous expression changes on their phylogeny. To do this, we treated the expression log ratio values as quantitative characters and used them to estimate the species-level phylogeny of the *Gossypium* allotetraploids using the contml program from the PHYLIP package (Felsenstein, 2005). The resulting 'homoeolog expression' phylogram (Fig. 1b) has a similar topology to the known phylogeny (note the polytomy at the base of Fig. 1b compared with Fig. 1a). The branch lengths found on this 'expression tree' are proportional to the levels of expression deviation from a common ancestor. From this representation, it is clear that *G. tomentosum* has experienced the greatest amount of total expression evolution. This is because *G. tomentosum* has a large number of A and D biases (Table 2) and, in addition, many of these biases are quite extreme, as indicated by the total branch length in Fig. 1b, which is a function of the total deviation from expression equivalence. Furthermore, *G. barbadense*, which has similar numbers of biased genes when compared with *G. tomentosum* (Table 2), has less overall deviation from its common ancestor with *G. darwinii* (a wild Galapagos Islands endemic) than might be expected. This effect reflects the fact that, although many *G. barbadense* homoeologs are expressed in a manner that is statistically biased, they do not deviate from equivalence to the degree found in *G. tomentosum*. The distribution of homoeolog expression levels for all species is shown in Fig. 2, which depicts, in histogram form, the expression log ratios for all 1383 genes. These histograms visually capture the significant differences in the level of deviation from equivalent expression in each of the species, with *G. tomentosum* having a broad profile relative to the other species, consistent with its high level of homoeolog expression divergence, and the F_1 hybrid having a narrow profile, consistent with its low deviation from equivalent genomic expression. Also evident, although perhaps subtle, is the overall D genome bias, which is evidenced by a greater density of values below zero than above. An additional dimension of this D genome bias is also revealed by the histograms, namely, that it is not caused, for example, by a large number of genes with an extreme D bias, but rather by an overall accumulation of many small D biases.

Global categorization of expression profiles and genomic dominance among allotetraploid cottons

Beyond the examination of homoeologous expression for 1383 genes, we also studied the overall duplicate gene expression for the five natural allopolyploids and the synthetic hybrid for 42 459 genes, using comparisons between each of these taxa and their A and D genome parents. The probes used to measure expression among these genes are generic with respect to the A and D genomes, meaning that they measure the cumulative output of both homoeologs, rather than homoeolog-specific expression as in the previous

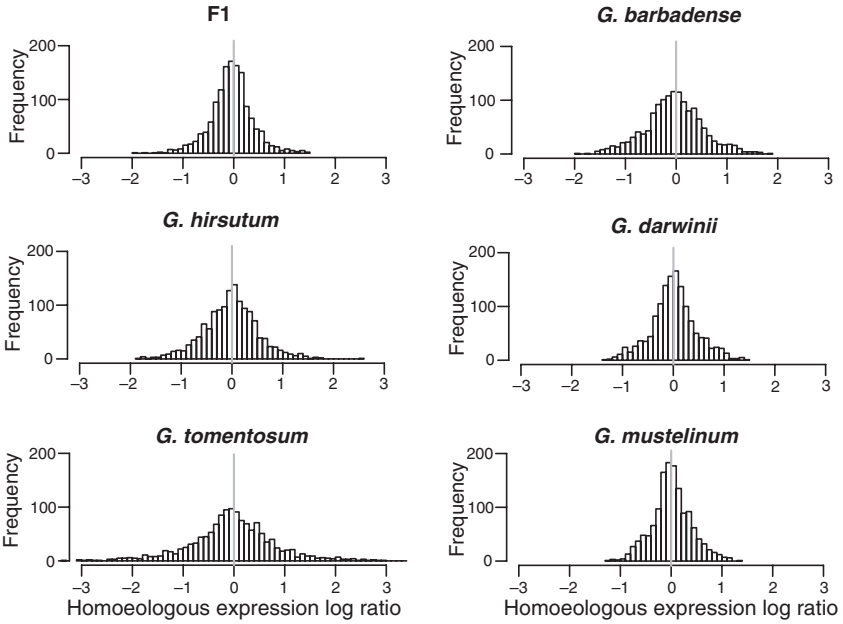

Fig. 2 Histograms of homoeologous expression among a synthetic F_1 hybrid and five *Gossypium* allotetraploids. All homoeolog expression values are expressed as log ratios [$\log_e(A_{probe}) - \log_e(D_{probe})$]; thus, positive values indicate greater A genome expression and negative values indicate greater D genome expression. A grey vertical bar on each histogram marks zero.

section. Within an allopolyploid, these generic probes can, however, be used to detect expression evolution in the form of nonadditive expression states (meaning that allotetraploid expression is not equivalent to the average expression of the parental species), such as parental dominance and transgressive up- or down-regulation (Wang *et al.*, 2006; Rapp *et al.*, 2009). Using this approach, Rapp *et al.* (2009) showed that this type of expression data can be parsed into 12 informative categories of expression evolution, to which they gave the Roman numeral designations seen across the top of Fig. 3. These include two forms of additive expression (I and XII; Fig. 3), which represent the null hypothesis, as well as genomic dominance (II, IV, IX and XI) and transgressive up- (V, VI and VIII) and down-regulation (III, VII and X). As used in Rapp *et al.* (2009) and here as well, the term genomic dominance refers to cases in which the expression state in the allopolyploid mimics that of one of its two diploid progenitors, irrespective of whether the direction is up-regulation or down-regulation of the A genome diploid relative to the D genome diploid. Transgressive expression is defined as statistically elevated or depressed expression relative to the two progenitor diploids.

For each of these 12 evolutionarily informative categories and for each species, we tallied the gene counts from among the genes assessed by our microarray, together with a tally of genes that showed statistically equivalent expression among the A and D genome parents and the allotetraploid or F_1 hybrid ('No Change'; Fig. 3). This analysis revealed that the levels of additive expression (I and XII) are relatively stable among all species. In addition, the amounts of A genome and D genome dominance (IV and IX vs II and XI, respectively) are also approximately equal among all five allopolyploids and with respect to the direction of dominance within species. By contrast, the F_1 hybrid displays about double the level of D genome expression dominance (II and XI) when compared with the reciprocal forms of A genome dominance (II and XI). An additional difference is that all of the natural allotetraploids show more transgressive up- (V, VI and VIII) and down-regulation (III, VII and X) than is observed in the F_1 hybrid, by approximately a factor of 10. Finally, within each of these categories, there is some variation between the allopolyploids, although this variation is smaller than that between any of the allopolyploids and the F_1 hybrid, and is probably constrained to an extent by a shared evolutionary history (Fig. 1a).

Discussion

The pace of expression evolution during polyploid formation, stabilization and speciation

Previous analyses in *Gossypium* have shown that genome merger, genome duplication and subsequent duplicate gene evolution each play roles in the alteration of homoeologous expression profiles (Flagel *et al.*, 2008; Chaudhary *et al.*, 2009a). These studies used *G. hirsutum* as the only allotetraploid representative, whereas, in the present study, we provide additional support for these findings by showing that all *Gossypium* allotetraploids have significant levels of homoeologous expression bias (much more so than does the F_1 hybrid; Table 2, Fig. 2). Moreover, in each of the five species, these biases favor the D genome. Because these characteristics are found throughout the allotetraploid phylogeny, they are inferred to have arisen: (1) after allopolyploid formation but before speciation; (2) recurrently after speciation in each allotetraploid lineage; or (3) a combina-

Counts of allotetraploid/parental generic expression patterns

Categories	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
	♀Allo♂	♀Allo♂	♀Allo♂	♀Allo♂	♀Allo♂	♀Allo♂	♀Allo♂	♀Allo♂	♀Allo♂	♀Allo♂	♀Allo♂	♀Allo♂
G. hirsutum (No change = 18 323)	1581	3240	621	3496	553	523	1855	1563	4444	497	4016	1747
G. barbadense (No change = 14 527)	2067	3750	1097	3453	781	803	3502	2371	3320	733	3902	2153
G. tomentosum (No change = 11 556)	2064	2927	1489	3578	1030	1290	4387	3666	3857	1036	3403	2176
G. mustelinum (No change = 14 793)	1827	3498	978	3601	825	781	3281	2371	3844	800	3863	2024
G. darwinii (No change = 12 561)	2069	3050	1274	4010	924	991	4056	3199	3958	937	3217	2213
F_1 (No change = 23 878)	1581	4888	248	2264	69	168	452	302	1951	60	4629	1951

Fig. 3 The 13 possible expression states in an F_1 hybrid and *Gossypium* allotetraploids relative to their model maternal (*G. arboreum*) and paternal (*G. ramondii*) progenitors. Each of the 12 states involving differential expression is labeled with a Roman numeral, using the same categorization as Rapp *et al.* (2009), and including a cartoon depiction, where maternal (♀; *G. arboreum*) and paternal (♂; *G. raimondii*) states are on the outer edges and the polyploid or F_1 hybrid (labelled 'Allo') state is in the middle. Expression values on the same horizontal line indicate statistically equivalent expression, whereas expression values on higher or lower horizontal lines represent statistically significant up- and down-regulation, respectively. For example, genes under Roman numeral I are those in which expression in the A genome diploid is significantly less than it is in the D genome diploid, with gene expression in the allopolyploid being significantly different from both but intermediate in value. The number of genes falling into a 13th category of gene expression, where there is no significance in any comparison among the two diploids and the allopolyploid, is shown by each species' name following 'No Change'.

tion of both, that is, with an immediate effect on allopolyploid formation followed by enhancement or elaboration during diversification in the subsequent 1–2 million years. The phylogenetic framework adopted here is illustrative in this respect.

The partitioning of gene expression evolution into its temporal components leads to the suggestion that these different components may entail different or at least complementary mechanisms. The first, involving rapid or instantaneous gene expression alteration as a consequence of genome merger and doubling, reflects the myriad novel interactions accompanying a biological reunion of two differentiated genomes into a common nucleus. The precise nature of these interactions is not known, but probably includes disruptions in gene dosage balance, stoichiometric changes resulting from differences in competition for transcription factors, differences in microRNA expression and a host of novel *cis* and *trans* interactions (Birchler and Reitia, this volume; Osborn *et al.*, 2003; Veitia, 2005; Veitia *et al.*, 2008). These saltational changes also probably involve genomic dominance, *sensu* Rapp *et al.* (2009), who demonstrated that gene expression in a synthetic *Gossypium* allopolyploid is strongly biased towards one of the two parental diploid genomes.

Superimposed on these rapid evolutionary responses to polyploidy are those that arise more slowly during the stabilization of the new polyploid genome and during evolution and speciation over much longer periods of time. Of particular interest are the striking changes with respect to the phenomenon of genomic dominance and the emergence of a high level of transgressive segregation (discussed in the next section), and also the continued elaboration of homoeolog bias that first becomes evident in the F_1 diploid hybrid (Table 2). That these changes continue to occur following speciation, in each allopolyploid lineage, is evidenced by the large number of genes exhibiting bias (Table 2) and the strikingly different rates of overall homoeolog expression evolution (Fig. 1b). Thus, the presence of duplicated genomes would seem to provide evolutionary opportunity and consequences immediately on polyploid formation and for millions of years thereafter. This, of course, is not a novel realization; indeed, it is one that has motivated many of the papers in this volume. Nonetheless, our results provide a novel dimension to this axiom, by demonstrating the temporal partitioning and phylogenetic perspective on global patterns of duplicate gene pair expression evolution.

A temporal component to the evolution of genomic dominance

In addition to the temporal components of expression evolution discussed above, our data reveal a surprising dimension to the newly described phenomenon of genomic dominance. Specifically, the synthetic F_1 hybrid used in this study and the synthetic AADD allopolyploid used by Rapp *et al.* (2009) both show strong evidence for genomic

dominance, whereby the D genome parental expression state is exhibited in strong preference over the A genome parental expression state. These data can be found in Fig. 3 of this paper for the F_1 hybrid and in figure 3 of Rapp *et al.* (2009) for the synthetic allotetraploid. Summarizing these data, the level of D dominance is as follows: category II = 4888 and 5719 and category XI = 4629 and 5257, for the F_1 hybrid used in this study and the synthetic allotetraploid used by Rapp *et al.* (2009), respectively. This is contrasted with A dominance, which includes category IV = 2264 and 663 and category IX = 1951 and 119, again for the F_1 hybrid and synthetic allotetraploid, respectively. This D dominance effect is also observed at the homoeolog level for the F_1 hybrid, as there are more than twice as many D genome biases as A genome biases (334 vs 153; Table 2).

A key point emerges from a comparison of the foregoing results with those observed in natural allopolyploids, which have had 1–2 million years to adjust to their polyploid condition. Specifically, the over-representation of the D genome bias is largely reversed among all five natural allopolyploids, both at the homoeolog level (Table 2) and among total gene expression profiles (Fig. 3). That is, over evolutionary time, the allotetraploids begin to assume roughly equivalent numbers of A and D dominant states. Interestingly, it is not the magnitude of genomic dominance that is altered by time, but its direction. That is to say, the levels of A (categories IV and IX; Fig. 3) and D (categories II and XI) dominance are nearly equivalent within each allotetraploid species, and both categories contain a large number of genes (*c.* 3000–4000), nearly equaling the magnitude of dominance found in the F_1 hybrid and synthetic allotetraploid from Rapp *et al.* (2009), where dominance was strongest only for the D genome. Thus, it would appear that the allopolyploid genomes have adjusted, during 1–2 million years of evolution, to more equally utilize the transcriptomes of the two co-resident genomes, although an appreciable level of D genome homoeolog bias remains evident in all five species (Table 2). It is possible that this residual D genome homoeolog bias is connected causally to the massive D dominance that arises following genome merger; to the extent that it is, it leads to the suggestion that homoeolog bias in other angiosperm allopolyploids may be predicted by the initial conditions established by genomic merger in the distant past (and which may be experimentally mirrored in many systems through the use of synthetic hybrids and allopolyploids).

A temporal component to the evolution of transgressive gene expression

As discussed above, the level of bias in genomic dominance has decreased in the 1–2 million years since the A and D genomes first became reunited. By contrast, transgressive up- and down-regulations are far more frequent among all extant allopolyploids than among the F_1 hybrid and synthetic allotetraploid used in Rapp *et al.* (2009). The values of transgressive expression from Rapp *et al.* (2009) are as follows: transgressive-up: V = 81, VI = 238, and VIII = 102; transgressive-down: III = 27, VII = 23 and X = 19; these values are strikingly lower than the values for the same expression categories in the five natural allotetraploids (Fig. 3). Moreover, the majority of the genes displaying transgressive expression patterns in the allotetraploids are found in the 'No Change' category in the F_1 hybrid (percentages range between *c.* 67 and 73% for the five allotetraploids), indicating that these transgressive states have probably evolved *de novo* in the allotetraploids from equal parental expression. From these results, we conclude that the instantaneous effect of genomic merger among the *Gossypium* A and D genomes is to create a significant level of D genome dominance, but not transgressive gene expression levels, regardless of ploidy level, whereas, over an evolutionary time-scale, all five natural allotetraploid species have alleviated the D genome control, but have evolved a large number of transgressive expression states. These findings suggest that the mechanism(s) underlying the high levels of transgressive expression within the natural allotetraploids differs from the instantaneous mechanisms that create D genome dominance. Because massive transgressive expression is only detected in the ancient allopolyploids, we speculate that long-term evolutionary processes, such as natural selection and *cis*- and *trans*-regulatory evolution, may play a role in their establishment.

Because the diploid A and D genome species used in this study are not the exact parents of the allotetraploids, we cannot say definitively that the differences described above are not the result of a different ancestry. However, a significant body of evidence indicates that *G. arboreum* and *G. raimondii* are the best extant models for the parents of allotetraploid cotton (reviewed in Wendel & Cronn, 2003), and that these diploid species are highly similar to the corresponding allopolyploid genomes at the sequence level (Senchina *et al.*, 2003; Grover *et al.*, 2004; Grover *et al.*, 2007). Therefore, it is likely that our temporal findings are genuine, and are unlikely to have arisen as an artifact of the discrepancies between the model diploid progenitors used in this study and the actual parents of the *Gossypium* allotetraploids.

Natural history and its effect on expression evolution in *Gossypium* allotetraploids

To our knowledge, this is the first analysis of the relative rate of expression evolution among homoeologs in plants. One of the more interesting aspects of the results is the high level of rate variation (Table 2, Figs 1b,3) among relatively closely related species. Among the five natural allotetraploids used in this study, two species were represented by elite culti-

vars from a domesticated background (*G. barbadense* cv. Pima S7 and *G. hirsutum* cv. Maxxa), whereas the other three species, *G. mustelinum*, *G. darwinii* and *G. tomentosum*, are wild, the last two being island endemics, and *G. mustelinum* restricted to a small native range in northeastern Brazil (Wendel *et al.*, 1994; Wendel & Cronn, 2003). Interestingly, although both domesticates show significant levels of homoeologous expression bias (Table 2 and Fig. 1b), neither are as strongly biased as the wild species *G. tomentosum*. This bias in *G. tomentosum* is striking in terms of its magnitude, involving 88% of all duplicate gene pairs studied (Table 2), and with respect to the fact that it reflects biased transcription in both directions, i.e. towards the A genome for 552 duplicates and towards the D genome for 666 duplicates. These two features together create a long branch in the 'expression phenogram' shown in Fig. 1b.

Among the domesticates, it is possible that some alteration in expression is the byproduct of artificial selection during domestication. However, our study focuses on petal tissues, whose phenotypes were not consciously under selection during domestication and subsequent crop improvement. We note that both domestication and island colonization entail genetic bottlenecks, events that may trigger the release of epigenetic variation (Rapp & Wendel, 2005), potentially contributing to the varied expression patterns and phenotypes found among *G. barbadense*, *G. hirsutum*, *G. tomentosum* and *G. darwinii*. Interestingly, however, *G. darwinii*, the other island endemic, has less biased expression patterns than does *G. tomentosum*, indicating that other variables are involved and that there may be an idiosyncratic nature to homoeologous expression evolution during speciation.

Finally, in the foregoing paragraphs, we have emphasized the differences among allotetraploids, although we note that there is also substantial conservation among biased genes. Overall, the intersection of A- and D-biased genes from among the five allotetraploids lies in the range 29–60% of the genes in each species, a value calculated first by finding the intersection of all A- or D-biased genes across all allotetraploid accessions (Table 2, last row), and dividing these values by the A- or D-biased genes found in each individual allotetraploid (e.g. 208 D-biased genes shared by all allotetraploids divided by 441 D-biased genes in *G. darwinii* reveals that *c.* 47% of all D biases found in *G. darwinii* are shared with all other allotetraploids, and therefore probably ancestral D biases). Viewing these data in this light shows that there is a considerable level of conservation that may, in part, stem from ancestral biases inherited and maintained by all species.

Acknowledgments

We gratefully acknowledge the National Research Initiative of the USDA Cooperative State Research, Education and Extension Service (2005-35301-15700 to J.F.W.) and National Science Foundation Plant Genome Research Program (0638418 to J.F.W.) for their support. We also thank Nathan Springer and Bob Stupar for their help in developing the Sequenom platform used in validating our microarray results, and the University of Minnesota BioMedical Genomics Center for processing all Sequenom assays. James McD. Stewart and David Stelly kindly generated and shared the F_1 hybrid used in this study. Finally, we thank Dan Nettleton for statistical guidance and two anonymous reviewers for their helpful comments.

References

Barker MS, Kane NC, Matvienko M, Kozik A, Michelmore RW, Knapp SJ, Rieseberg LH. 2008. Multiple paleopolyploidizations during the evolution of the Compositae reveal parallel patterns of duplicate gene retention after millions of years. *Molecular Biology and Evolution* **25**: 2445–2455.

Bottley A, Koebner RM. 2008. Variation for homoeologous gene silencing in hexaploid wheat. *Plant Journal* **56**: 297–302.

Bottley A, Xia GM, Koebner RMD. 2006. Homoeologous gene silencing in hexaploid wheat. *Plant Journal* **47**: 897–906.

Buggs RJA, Doust AN, Tate JA, Koh J, Soltis K, Feltus FA, Paterson AH, Soltis PS, Soltis DE. 2009. Gene loss and silencing in *Tragopogon miscellus* (Asteraceae): comparison of natural and synthetic allotetraploids. *Heredity* **103**: 73–81.

Chaudhary B, Hovav R, Rapp R, Verma N, Udall JA, Wendel JF. 2008. Global analysis of gene expression in cotton fibers from wild and domesticated *Gossypium barbadense*. *Evolution and Development* **10**: 567–582.

Chaudhary B, Flagel L, Stupar RM, Udall JA, Verma N, Springer NM, Wendel JF. 2009a. Reciprocal silencing, transcriptional bias and functional divergence of homoeologs in polyploid cotton (*Gossypium*). *Genetics* **182**: 503–517.

Chaudhary B, Hovav R, Flagel L, Mittler R, Wendel JF. 2009b. Parallel expression evolution of oxidative stress-related genes in fiber from wild and domesticated diploid and polyploid cotton (*Gossypium*). *BMC Genomics* **10**: 378.

Chen ZJ. 2007. Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annual Review of Plant Biology* **58**: 377–406.

Comai L. 2005. The advantages and disadvantages of being polyploid. *Nature Reviews Genetics* **6**: 836–846.

Conant GC, Wolfe KH. 2008. Turning a hobby into a job: how duplicated genes find new functions. *Nature Reviews Genetics* **9**: 938–950.

Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, Wendel JF. 2008. Evolutionary genetics of genome merger and doubling in plants. *Annual Review of Genetics* **42**: 443–461.

Felsenstein J. 2005. *PHYLIP (Phylogeny Inference Package) version 3.6*. Distributed by the author. Seattle, WA, USA: Department of Genome Sciences, University of Washington.

Flagel L, Udall JA, Nettleton D, Wendel JF. 2008. Duplicate gene expression in allopolyploid *Gossypium* reveals two temporally distinct phases of expression evolution. *BMC Biology* **6**: 16.

Force A, Lynch M, Pickett FB, Amores A, Yan Y-l, Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**: 1531–1545.

Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC. 2007. Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. *Plant Cell* **19**: 3403–3417.

Grover CE, Kim H, Wing RA, Paterson AH, Wendel JF. 2004. Incongruent patterns of local and global genome size evolution in cotton. *Genome Research* **14**: 1474–1482.

Grover CE, Kim H, Wing RA, Paterson AH, Wendel JF. 2007. Microcolinearity and genome evolution in the *AdhA* region of diploid and polyploid cotton (*Gossypium*). *Plant Journal* **50**: 995–1006.

Hegarty MJ, Barker GL, Wilson ID, Abbott RJ, Edwards KJ, Hiscock SJ. 2006. Transcriptome shock after interspecific hybridization in *Senecio* is ameliorated by genome duplication. *Current Biology* **16**: 1652–1659.

Hovav R, Udall JA, Chaudhary B, Flagel L, Rapp R, Wendel JF. 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**: 6191–6195.

Krapovickas A, Seijo G. 2008. *Gossypium ekmanianum* (Malvaceae), algodón silvestre de la República Dominicana. *Bonplandia* **17**: 53–63.

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

Nettleton D, Hwang JTG, Caldo RA, Wise RP. 2006. Estimating the number of true null hypotheses from a histogram of *p* values. *Journal of Agricultural, Biological, and Environmental Statistics* **11**: 337–356.

Ni Z, Kim ED, Ha M, Lackey E, Liu J, Zhang Y, Sun Q, Chen ZJ. 2009. Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. *Nature* **457**: 327–331.

Ohno S. 1970. *Evolution by gene duplication*. New York, NY, USA: Springer-Verlag.

Osborn TC, Pires JC, Birchler JA, Auger DL, Chen ZJ, Lee H-S, Comai L, Madlung A, Doerge RW, Colot V *et al.* 2003. Understanding mechanisms of novel gene expression in polyploids. *Trends in Genetics* **19**: 141–147.

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

Pires JC, Zhao J, Schranz ME, Leon EJ, Quijada PA, Lukens LN, Osborn TC. 2004. Flowering time divergence and genomic rearrangements in resynthesized *Brassica* polyploids (Brassicaceae). *Biological Journal of the Linnean Society* **82**: 675–688.

Pumphrey M, Bai J, Laudencia-Chingcuanco D, Anderson O, Gill BS. 2009. Nonadditive expression of homoeologous genes is established upon polyploidization in hexaploid wheat. *Genetics* **181**: 1147–1157.

Rabinowicz PD, Citek R, Budiman MA, Nunberg A, Bedell JA, Lakey N, O'Shaughnessy AL, Nascimento LU, McCombie WR, Martienssen RA. 2005. Differential methylation of genes and repeats in land plants. *Genome Research* **15**: 1431–1440.

Rapp RA, Wendel JF. 2005. Epigenetics and plant evolution. *New Phytologist* **168**: 81–91.

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

Senchina DS, Alvarez I, Cronn RC, Liu B, Rong J, Noyes RD, Paterson AH, Wing RA, Wilkins TA, Wendel JF. 2003. Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. *Molecular Biology and Evolution* **20**: 633–643.

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

Small RL, Ryburn JA, Cronn RC, Seelanan T, Wendel JF. 1998. The tortoise and the hare: choosing between noncoding plastome and nuclear *Adh* sequences for phylogeny reconstruction in a recently diverged plant group. *American Journal of Botany* **85**: 1301–1315.

Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, dePamphilis CW, Wall PK, Soltis PS. 2009. Polyploidy and angiosperm diversification. *American Journal of Botany* **96**: 336–348.

Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences, USA* **100**: 9440–9445.

Stupar RM, Springer NM. 2006. *Cis*-transcriptional variation in maize inbred lines B73 and Mo17 leads to additive expression patterns in the F_1 hybrid. *Genetics* **173**: 2199–2210.

Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. 2008. Synteny and collinearity in plant genomes. *Science* **320**: 486–488.

Tate JA, Ni Z, Scheen AC, Koh J, Gilbert CA, Lefkowitz D, Chen ZJ, Soltis PS, Soltis DE. 2006. Evolution and expression of homeologous loci in *Tragopogon miscellus* (Asteraceae), a recent and reciprocally formed allopolyploid. *Genetics* **173**: 1599–1611.

Tate JA, Joshi P, Soltis KA, Soltis PS, Soltis DE. 2009. On the road to diploidization? Homoeolog loss in independently formed populations of the allopolyploid *Tragopogon miscellus* (Asteraceae). *BMC Plant Biology* **9**: 80.

Udall JA, Swanson JM, Haller K, Rapp RA, Sparks ME, Hatfield J, Yu Y, Wu Y, Dowd C, Arpat AB *et al.* 2006a. A global assembly of cotton ESTs. *Genome Research* **16**: 441–450.

Udall JA, Swanson JM, Nettleton D, Percifield RJ, Wendel JF. 2006b. A novel approach for characterizing expression levels of genes duplicated by polyploidy. *Genetics* **173**: 1823–1827.

Veitia RA. 2005. Paralogs in polyploids: one for all and all for one? *Plant Cell* **17**: 4–11.

Veitia RA, Bottani S, Birchler JA. 2008. Cellular reactions to gene dosage imbalance: genomic, transcriptomic and proteomic effects. *Trends in Genetics* **24**: 390–397.

Wan C, Wilkins T. 1994. A modified hot borate method significantly enhances the yield of high-quality RNA from cotton (*Gossypium hirsutum* L.). *Analytical Biochemistry* **223**: 7–12.

Wang J, Tian L, Lee H-S, Wei NE, Jiang H, Watson B, Madlung A, Osborn TC, Doerge RW, Comai L *et al.* 2006. Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. *Genetics* **172**: 507–517.

Wendel JF. 2000. Genome evolution in polyploids. *Plant Molecular Biology* **42**: 225–249.

Wendel JF, Cronn R. 2003. Polyploidy and the evolutionary history of cotton. *Advances in Agronomy* **78**: 139–186.

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

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Validation results for four natural allotetraploid species.

Table S1 A and D genome biases and equivalent contribution to the transcriptome for additional *q*-value thresholds

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.