

Homoeologous nonreciprocal recombination in polyploid cotton

Armel Salmon¹, Lex Flagel¹, Bao Ying^{1,2}, Joshua A. Udall³ and Jonathan F. Wendel¹

¹Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA 50011, USA; ²College of Life Science, Qufu Normal University, Qufu, Shandong 273165, China; ³Department of Plant and Wildlife Sciences, Brigham Young University, Provo, UT 84602, USA

Summary

Author for correspondence:

Jonathan F. Wendel

Tel: +1 515 294 7172

Email: jfw@iastate.edu

Received: 26 August 2009

Accepted: 6 October 2009

New Phytologist (2010) **186**: 123–134

doi: 10.1111/j.1469-8137.2009.03093.x

Key words: allopolyploid speciation, cotton, expressed sequence tag (EST) assembly, gene conversion, *Gossypium*, homoeologous recombination.

- Polyploid formation and processes that create partial genomic duplication generate redundant genomic information, whose fate is of particular interest to evolutionary biologists. Different processes can lead to diversification among duplicate genes, which may be counterbalanced by mechanisms that retard divergence, including gene conversion via nonreciprocal homoeologous exchange.
- Here, we used genomic resources in diploid and allopolyploid cotton (*Gossypium*) to detect homoeologous single nucleotide polymorphisms provided by expressed sequence tags from *G. arboreum* (A genome), *G. raimondii* (D genome) and *G. hirsutum* (AD genome), allowing us to identify homoeo-single nucleotide polymorphism patterns indicative of potential homoeologous exchanges.
- We estimated the proportion of contigs in *G. hirsutum* that have experienced nonreciprocal homoeologous exchanges since the origin of polyploid cotton 1–2 million years ago (Mya) to be between 1.8% and 1.9%. To address the question of when the intergenomic exchange occurred, we assayed six of the genes affected by homoeo-recombination in all five *Gossypium* allopolyploids using a phylogenetic approach.
- This analysis revealed that nonreciprocal homoeologous exchanges have occurred throughout polyploid divergence and speciation, as opposed to saltationally with polyploid formation. In addition, some genomic regions show multiple patterns of homoeologous recombination among species.

Introduction

Plant genomes offer unique opportunities to study the coexistence of duplicated gene copies within the same nucleus. These duplicated copies originate via several mechanisms that generate tandem and dispersed duplicates, including replication errors, retropositioning and other transposable element-mediated processes, and, perhaps most importantly, polyploidy, which causes the simultaneous duplication of all genes. These duplications set in motion a spectrum of evolutionary responses, caused by the opportunities afforded by genic redundancy and the accompanying alteration of functional and selective constraints on one or both duplicated genes. Thus, there has long been interest in the creative aspect of gene duplication ('neofunctionalization'), with a focus on its potential role in the generation of evolutionary novelty and hence its relevance to adaptation and diversification

(Stephens, 1951; Ohno, 1970; Doyle *et al.*, 2008). More recently, a broadened perspective has emerged on the fuller set of evolutionary outcomes of gene duplication, including pseudogene formation and various kinds of sub- and neofunctionalization (Lynch & Conery, 2000; Conant & Wolfe, 2008; Flagel & Wendel, 2009). These diversification processes may be counterbalanced by mechanisms that retard the divergence of duplicates, including preservation or maintenance of redundant gene function as a result of strong selection, and gene conversion mediated by nonreciprocal homoeologous exchange. Although these topics have received relatively less attention than those that generate divergence between duplicates, they are particularly amenable to study in relatively recently formed allopolyploid species. In such cases, two diverged genomes that evolved independently in diploid lineages are reunited into a common nucleus at the time of polyploid formation, thus providing a natural context for direct compari-

sons of sequence, expression and function among genes in diploid antecedents and their allopolyploid derivatives.

In principle, recombination between homoeologous loci may arise during meiosis or mitosis in cell lineages destined to become germinal. Recent analysis of segregating populations using molecular markers in *Brassica napus* highlighted the presence of homoeologous exchanges (Parkin *et al.*, 1995; Udall *et al.*, 2005), leading to the observation in progenies of nonreciprocal transpositions between homoeologous genomes (discussed in Nicolas *et al.*, 2008). Of particular relevance here are the two mechanisms leading to nonreciprocal exchanges of homologous chromatids: via crossovers (COs) and the subsequent segregation of one parental and one recombinant chromatid, or via noncrossovers (NCOs, previously named gene conversion events). The most commonly invoked mechanism leading to NCOs is the synthesis-dependent strand annealing pathway, rather than the double-strand break repair model (Szostak *et al.*, 1983) after a double-strand break event (Paques & Haber, 1999; Hollingsworth & Brill, 2004), involving the displacement of one homologous chromatid with the other, leading to the synthesis of homologous DNA (reviewed in de Massy, 2003; Chen *et al.*, 2007; Mezard *et al.*, 2007; Martinez-Perez & Colaiacovo, 2009). In addition to simple gene conversion between homologs, gene conversion may also include the conversion of paralogs (in the case of homogenized gene families; Teshima & Innan, 2004), or even a form of diversifying conversion when a short segment of a novel allele or pseudogene (Chen *et al.*, 2007) is used as a conversion template, potentially leading to a chimeric product. These same processes, classically modeled to describe the segregation of homologous chromatids, are applicable to *homoeologous* sequences in allopolyploids, and can both lead to the observation in their derivative progenies of nonreciprocal homoeologous recombination (NRHR).

Cotton (*Gossypium*) allopolyploids (AD genomes; $2n = 52$) originated 1.5 million years ago (Mya) (Senchina *et al.*, 2003; Wendel & Cronn, 2003) from hybridization between an A genome ($2n = 26$), African species much like modern *G. arboreum* and *G. herbaceum*, and a D genome ($2n = 26$), American species similar to modern *G. raimondii*. Presently, five AD-genome polyploids are widely recognized, although a sixth species has been proposed recently (Krapovickas & Seijo, 2008). These species (and their genomic designations) include *G. hirsutum* (AD1), or 'Upland cotton', accounting for 90% of the global cotton produced (Wendel & Cronn, 2003), *G. barbadense* (AD2), also cultivated and commonly referred to as 'Pima' or 'Egyptian' cotton, and three other exclusively wild polyploid species endemic to coastal and island habitats: *G. tomentosum* (AD3), *G. mustelinum* (AD4) and *G. darwinii* (AD5). The presence of nucleotide synapomorphies (shared derived character states) among these five allopolyploid species indi-

cates that it is likely that they originated from a single polyploidization event (Small *et al.*, 1998; Small & Wendel, 1999).

The fate of duplicated genes has previously been studied in allopolyploid cotton, but this was before the recent explosion in the availability of expressed sequence tags (ESTs) and other sequence information. Cronn *et al.* (1999) used PCR and the sequencing of isolated clones to study homoeologous copies of 16 low-copy-number nuclear genes. Phylogenetic analyses indicated that, for this limited sampling of the genome, single- or low-copy-number genes were evolving independently. By contrast, homoeologous interactions were described for rDNA internal transcribed spacer sequences (Wendel *et al.*, 1995), which were inferred to have experienced bidirectional concerted evolution, meaning that four polyploid species exhibited only D-like copies, whereas one (*G. mustelinum*) contained mostly or exclusively A-like copies. Such homoeologous interactions among rDNA arrays have since been documented in a variety of other polyploid genera, e.g. *Nicotiana* (Kovarik *et al.*, 2004, 2008) and *Tragopogon* (Kovarik *et al.*, 2005).

Our goal in the present study was to utilize a vastly expanded EST database (Udall *et al.*, 2006a; J. A. Udall, unpublished) to assess the extent of NRHR in allopolyploid *Gossypium*. These EST resources were developed using cDNA libraries from the two diploid progenitor genomes (A, D) of allopolyploid cotton and from multiple cDNA libraries derived from diverse tissues of allopolyploid cotton, thereby allowing homoeologous single nucleotide polymorphisms (homoeo-SNPs) to be identified in the allotetraploid genome and assigned to the appropriate diploid parent of origin (Udall *et al.*, 2006b). By analyzing the homoeo-SNP patterns in the allopolyploid ESTs and contigs, putative homoeologous recombined regions were detected, whereby ESTs in the allopolyploid apparently derived from one parental EST displayed at least one homoeo-SNP signature from the other parent. An NRHR was tentatively scored when one homoeologous sequence displayed such recombination, whereas the other homoeolog displayed a parental copy. To validate these putative homoeo-recombination events, we cloned and resequenced a subset of them, revealing, in the process, several of the artifacts that could lead to the erroneous inference of homoeologous gene conversion, such as autapomorphies (derived character states unique to either one of the parents or the polyploids) and EST assembly errors. To ensure that recombinant ESTs were not caused by post-transcriptional events such as co-splicing of exons from homoeologous A- and D-genome mRNA copies, a subset of primers was designed to amplify both exonic and intronic regions. Finally, we used the phylogenetic context provided by the five natural allopolyploid species to address the question of whether gene conversion is a process that is most probable during the early stages of

polyploid stabilization or, instead, is a phenomenon that arises more slowly on an evolutionary timescale.

Materials and Methods

EST assemblies

The detection of recombination events between homoeologous loci was facilitated using the published cotton EST assembly (Udall *et al.*, 2006a) to which 1 750 000 new EST reads were added using a next-generation sequencing method (Roche-454 sequencing). This EST assembly is deep and extensive in terms of presumed gene content coverage, consisting of aligned sequences from *G. arboreum* (A genome), *G. raimondii* (D genome) and *G. hirsutum* (AD polyploid genome), generated from multiple cDNA libraries extracted from different organs, at different developmental stages, and in response to different biotic and abiotic stresses. In the assembly process, ESTs from both diploids and from the allotetraploid were included, creating a multispecies assembly, which allows for the detection of homoeo-SNPs. This new cotton EST database contains 19 874 contigs assembling A-, D- and AD-genome ESTs from which homoeo-SNPs can be detected. All EST sequences and the assembly are publicly available at <http://www.agcol.arizona.edu/cgi-bin/pave/Cotton/index.cgi>.

Detection of homoeologous recombinants

A Python-Biopython custom script was used to detect homoeologous recombinants. This program traverses alignments of putatively orthologous EST sequences from the diploid A and D genomes to detect diagnostic SNPs. SNPs in the diploids that have been transmitted from the time of polyploidization and have remained unaltered since polyploid formation comprise diagnostic homoeo-SNPs in the allopolyploid. We used a conservative approach to identify genome-specific SNPs by requiring that all available A- or D-genome diploid sequences be in complete agreement with regard to all diagnostic nucleotide positions. Inferred genome-specific SNPs were compared with ESTs from allo-

polyploid cotton to infer the fate and linkage of homoeo-SNPs. When an EST from an allopolyploid displayed both A- and D-genome homoeo-SNPs, it was flagged as a potential homoeologous recombinant. An example of the inference of homoeologous recombinants is illustrated in Fig. 1, which, for clarity, shows both types of unaltered homoeologs (At, Dt), as well as one that might be expected from a recombination event between homoeologous loci. A region presenting a conserved homoeologous sequence (At or Dt) and a homoeologous recombinant event was inferred as revealing an NRHR.

In addition to estimating the prevalence of NRHR throughout the cotton EST assembly, a second, more conservative script was used that considered the possible effects of autapomorphic substitutions (Fig. 1), in the 1–2 Myr since polyploid formation, as these polymorphisms can generate false positives with respect to NRHR estimation. The application of this script led to a decreased proportion of inferred NRHR events.

Validation of bioinformatically inferred homoeologous recombination events

We used traditional Sanger sequencing to verify putative NRHR events from genomic DNA. This work was viewed as essential in order to estimate the effect of EST assembly and other possible artifacts on inferences of recombination events between homoeologous loci. Accordingly, we selected a subset of these putative events and designed amplification primers for PCR, followed by amplicon cloning and sequencing.

Plant material

We sampled putative NRHR events from the species used in the EST sequencing, and representatives of the four other AD-genome polyploid species. These lines, which included *G. arboreum* (A genome), *G. raimondii* (D genome), *G. hirsutum* cv. Acala maxx (AD1 genome), *G. barbadense* cv. Pima S6 (AD2 genome), *G. tomentosum* accession WT936 (AD3 genome), *G. mustelinum* accession 15C (AD4



Fig. 1 Illustration of the detection of homoeologous single nucleotide polymorphisms (homoeo-SNPs) and nonreciprocal homoeologous recombination (NRHR) in diploid and allopolyploid cotton. Genome designations and sequences without gene conversion are as follows: A, *Gossypium arboreum*; D, *G. raimondii*; At, *G. hirsutum*, A-genome homoeolog; Dt, *G. hirsutum*, D-genome homoeolog; NRHRt, putative gene conversion event in *G. hirsutum*. The last sequence exhibits homoeo-SNPs corresponding to the A genome in the first and last SNP positions, but to the D genome in the middle SNP position. This is inferred to be an A homoeolog that has undergone NRHR in its middle section by the D homoeolog. 1, homoeo-SNP; 2, parental autoapomorphy; 3, allopolyploid autoapomorphy.

genome) and *G. darwinii* accession PW45 (AD5 genome), were grown in the Pohl Conservatory at Iowa State University. In addition, for the purposes of phylogenetic reconstruction, we included the outgroup species *Gossypioides kirkii* (Seelanan *et al.*, 1997). Leaf punches were collected and DNA was extracted following a CTAB protocol (Doyle, 1991). The concentration of the DNA solutions was measured and normalized to 50 ng μl^{-1} .

Primer design, cloning and sequencing

Our goal was to design primers that would amplify genomic DNA from genes corresponding to EST sequences in different *Gossypium* species, and that would include either just exons or both exons and introns from regions displaying a putative homoeologous recombinant in *G. hirsutum*. To estimate exon–intron boundaries, we blasted our EST and contig sequences against the *Arabidopsis* genome (TAIR8 genomic sequences release, <http://www.arabidopsis.org/>), and aligned the cotton transcriptomic sequence to the closest *Arabidopsis* genomic sequence using Spidey (<http://www.ncbi.nlm.nih.gov/IEB/Research/Ostell/Spidey/index.html>). Gaps in the resulting alignments were inferred to correspond to putative introns. Primers were designed employing Primer3 v. 1.1.4 (Rozen & Skaletsky, 2000), using default parameters, and were selected using a script to find primer pairs that allowed amplification in both the A and D genomes, either within a single exon or between two or more exons.

Each PCR contained 0.5 μM of each primer, 200 μM of deoxynucleoside triphosphates, 1.5 mM of MgCl_2 , 1 \times *Taq* DNA polymerase buffer (Invitrogen), 1 unit of *Taq* DNA polymerase (Invitrogen) and 50 ng of DNA, in a final volume of 20 μl . To ensure thermocycling conditions that limit PCR recombination events (Judo *et al.*, 1998; Cronn *et al.*, 2002), we used a limited number of cycles (20–25) and long elongation times. The amplification program was 94°C for 2 min, followed by 20–25 cycles at 94°C for 30 s, 60°C for 30 s and 72°C for 1 min 30 s, and a final extension at 72°C for 5 min. The primer pairs selected for each targeted region are presented in Table S1 (see Supporting Information).

Amplified products were visualized following agarose gel electrophoresis. Bands of the expected size were excised from the gel, purified using a Qiaquick gel extraction kit (Qiagen, Hilden, Germany) and cloned using the TOPO TA cloning kit (Invitrogen). From the 20 regions of interest displaying a putative recombination between homoeologs that were selected for validation, six were selected and cloned to check both the A-genome and D-genome EST sequence quality and to obtain parental intronic sequences. These six regions were also cloned from the four other AD-polyplloid genomes (AD2, AD3, AD4 and AD5) and the outgroup *Gossypioides kirkii*. For each region of interest and

in each species in which a gene was cloned, 10–12 clones were sequenced by Sanger sequencing at the Iowa State University core sequencing facility.

Sequence analysis

Sequencing chromatograms were checked visually and alignments were processed using ClustalW v.2 (Larkin *et al.*, 2007). Each alignment and each SNP were verified visually. Open reading frames (ORFs) were detected using the ORF-Finder program (<http://www.ncbi.nlm.nih.gov/projects/gorf/>). Alignment editing was made using Jalview (Waterhouse *et al.*, 2009). For each species, we required at least two independent sequences to remove potential PCR recombination artifacts (Judo *et al.*, 1998; Cronn *et al.*, 2002). Regions displaying recombinants were blasted against genomic databases, and separately to the *Arabidopsis* genome database (TAIR9 cDNA sequences release), to identify putative gene type and function.

Results

Detection of nonreciprocal homoeologous recombination events using cotton homoeologous ESTs

The search for homoeologous recombinants allowed us to detect 485 putative events (disallowing parental autapomorphies) from the 19 874 contigs displaying A-, D- and AD-genome ESTs in the Cotton32 assembly. Of these 485 gene conversion candidates, 341 exhibited a single recombination breakpoint (which may be caused by either homoeologous CO and subsequent segregation of one parental copy and one recombinant, or NCO – gene conversion events), whereas 100 showed two recombination breakpoints, and 44 showed a pattern of three or more recombination breakpoints (Table 1). As our inferences rely on the use of EST assemblies, we may underestimate the actual number of recombination events found at the DNA level. The size range of the recombined regions displaying two recombina-

Table 1 Number of homoeologous recombinant copies detected per contig and patterns of recombination

	Number of homoeologous recombinant copies per contig				
	1	2	3	4	Σ
Number of recombination breakpoint(s)					
1	270	62	8	1	341
2	67	26	6	1	100
≥ 3	19	20	4	1	44
Σ	356	108	18	3	485

1, single; 2, double; ≥ 3 , three or more recombination breakpoints.

tion breakpoints ranged between 1 and 698 nucleotides in length ($N_{50} < 50$ bp). By considering a total of 356 contigs with only one recombinant copy (either D genome, A recombined copies or the reciprocal situation) in the AD1 genome (Table 1), we could conservatively extrapolate that NRHR had modified 1.8% of the cotton transcriptome. The 129 other candidate regions represented patterns for which 74 corresponded to reciprocal homoeologous recombinants (0.4%), 44 to nonreciprocal homoeologous recombinants for both homoeologous copies (0.2%), in which the two homoeologous copies displayed different recombination patterns, and five to recombinants from different regions within a contig (0.02%).

As EST assembly may greatly influence our approach to the detection of NRHR, we chose to calculate the proportion of NRHR events following allopolyploidy using a second more conservative assembly. Analysis of this assembly should minimize the number of EST assembly errors (e.g. paralogous copies assembled in the same contig, leading to the detection of a spurious homoeologous exchange). This reassembled EST database (Cotton41 assembly) contained 17 386 contigs from which homoeo-SNP signatures were inferred. Using this comparative approach, we estimated that the proportion of NRHRs in *G. hirsutum* (disallowing parental autapomorphies and taking only into consideration the contigs displaying nonreciprocal homoeologous exchanges affecting one homoeolog) was from 1.8% (357/19 874) to 1.9% (335/17 386). The overall proportion of homoeologous exchanges in the cotton transcriptome (considering all the contigs displaying recombinant ESTs) was from 2.4% (486/19 874) to 3.5% (608/17 386).

Validation of putative nonreciprocal homoeologous recombination events

Because bioinformatic inferences of homoeologous exchanges are subject to several possible sources of error (e.g. assembly artifacts, paralogy, sequencing error, autapomorphic homoplasy, RNA editing), it is important to validate at least some of the suspected cases using an independent method. This validation process has value not only in refuting or verifying putative intergenomic recombination, but also in shedding light on the quality of the assembly and the sequence alignments. Here, we chose to validate putative NRHRs and parental autapomorphies for 20 regions (contigs), selected on the basis of several criteria. These criteria included the number of ESTs (both from the diploids and the polyploid) included in the alignment that provided homoeo-SNP evidence, and the number and distribution of putatively converted homoeo-SNPs. The selected contigs (Table 2) ranged in length from 900 to 2203 nucleotides, and each displayed an ORF. The BLAST results presented in Table 2 correspond to alignments with

Arabidopsis thaliana, from which the best functional annotation is available. These contigs corresponded to various kinds of genes, some (Cotton16_07872_01, Cotton16_00080_04, Cotton16_14915_01, Cotton16_00255_03) displaying best homology to members of larger gene families (α -tubulin, β -tubulin, ankyrins and strubbeling-receptors, respectively), and others having homologies to single- or low-copy-number genes in *Arabidopsis*. The primer designing strategy used allowed the amplification of intronic regions for 11 of the 20 targeted genomic regions, one containing more than one intron (Cotton16_00255_03, with three introns).

For each sampled genomic region, both A- and D-genome homoeologous copies (AD1-A and AD1-D) were recovered from *G. hirsutum*. Inclusion into our final validated sequence dataset required that both homoeologous copies be detected, and that each be present at least twice. Cases in which a sequence was obtained only once were considered to be potential PCR recombination or sequencing artifacts, and were removed from the analysis.

Based on these *de novo* genomic sequences, the bioinformatically detected NRHR events were confirmed for 14 of the 20 regions analyzed (Table 3). Two of the remaining six contigs (Cotton16_01545_01 and Cotton28_24190) were not validated as having experienced a homoeologous recombination, but, instead, were shown to be mistaken inferences arising from the presence of paralogous or highly pseudogenized copies in the AD1 genome that were erroneously stitched together during EST assembly (these contigs showed two clearly distinct pools of sequences). The remaining four contigs (Cotton16_00047_02, Cotton16_01765_02, Cotton16_02054_02, Cotton28_21020) did not agree with the sequence of their respective contig, and may represent situations in which it was not possible to rule out autapomorphies in the diploid parents after polyploid formation *c.* 1.5 Mya. That is, a homoeo-SNP between modern A- and D-genome diploids may, in principle, reflect true homoeo-SNPs between the A and D genomes at the time of polyploid formation. These cases could be used as indicators or markers of NRHR in modern allopolyploid cotton. Alternatively, this homoeo-SNP may reflect an autapomorphic substitution after polyploid formation (Fig. 1), leading to an erroneous inference of intergenomic recombination (a false positive). Our detection method discriminated between these two types of SNP evidence.

Intronic sequences were obtained for six regions displaying confirmed recombination events between homoeologous copies (Cotton16_00255_03, Cotton16_01609_01, Cotton16_01955_02, Cotton16_10388_01, Cotton16_14915_01 and Cotton16_34881_01). The intron and intron–exon data provided two novel insights into illegitimate recombination within polyploid cotton. First, they confirmed that these events happened at the DNA level,

Table 2 Description of the regions selected to validate putative nonreciprocal homoeologous recombination (NRHR)

Contig	Length (nt)	ORF (start-end)	Intron start	Intron length	Start point (FP)	End point (RP)	BLAST results against <i>Arabidopsis</i> cDNAs			E value	
							Length cDNA	Symbol	Gene		
Cotton16_00047_02	1467	79–1248	–	–	966	1381	437	AT5G13930.1	CHS	Naringenin-chalcone synthase	0.0
Cotton16_00080_04	1607	64–1407	729	90	474	874	420	AT5G23860.1	TUB8	Tubulin β 8	0.0
Cotton16_00255_03	1219	220–1215	322–462–534	425–93–528	305	553	268	AT4G03390.1	SRF3	Strubelling-receptor family 3	6E–110
Cotton16_01545_01	972	400–969	–	–	126	546	421	AT1G64142.1	CPuORF23	Conserved peptide upstream open reading frame 23	8E–21
Cotton16_01609_01	1303	82–1299	532	80	385	884	500	AT2G26890.1	GRV2	Gravistropism defective 2/heat shock protein binding	0.0
Cotton16_01666_01	985	99–758	–	–	696	945	250	AT2G33120.1	SAR1	Synaptobrevin-related protein 1	5E–115
Cotton16_01765_02	1816	277–1629	–	–	1218	1647	430	AT3G49260.1	lqd21	lqd21 (IQ-domain 21); calmodulin binding	5E–41
Cotton16_01955_02	1001	51–999	216	341	140	623	484	AT1G80350.1	ERH3	Ectopic root hair 3	8E–137
Cotton16_02054_02	1082	39–1016	–	–	470	955	486	AT3G16150.1	–	L-asparaginase, putative	0.0
Cotton16_07872_01	1581	94–1446	–	–	872	1339	468	AT1G04820.1	TUA4	Tubulin α 4	0.0
Cotton16_10388_01	900	86–899	392	132	280	466	187	AT1G31780.1	–	Unknown expressed protein	3E–144
Cotton16_13240_01	1543	90–1335	–	–	1068	1552	485	AT1G79340.1	AtMCA4	Metacaspase 4; cysteine-type peptidase	3E–157
Cotton16_13273_01	2203	68–1999	1711	288	1567	1981	415	AT5G03340.1	CDC48	Cell division cycle protein 48, putative	0.0
Cotton16_14915_01	1222	149–1135	825	335	546	1000	455	AT5G65860.1	–	Ankyrin repeat family protein	4E–172
Cotton16_27501_01	1912	621–1649	–	–	1002	1278	277	AT4G28300.1	–	Hydroxyproline-rich glycoprotein family protein	2E–43
Cotton16_34868_01	1559	146–1240	962	100	939	1432	494	AT4G34160.1	CYCD3;1	Cyclin-dependent protein kinase regulator	2E–91
Cotton16_34881_01	1183	191–889	479	71	192	626	435	AT5G24890.1	–	Unknown protein	2E–28
Cotton28_21020	2171	239–790	1674	241	1637	2133	497	AT1G67840.1	CSK	Chloroplast sensor kinase	1E–172
Cotton28_21491	1039	22–630	–	–	434	909	476	AT3G26935.1	–	Zinc finger (DHC type) family protein	2E–78
Cotton28_24190	1408	1637–1903	1043	360	979	1467	489	AT2G20890.1	PSB29	Photosystem II binding	3E–60

FP, forward primer; ORF, open reading frame; RP, reverse primer.

Table 3 Description of the sequences cloned from regions displaying putative homoeologous recombination events in *Gossypium hirsutum*

Contig	Genome of origin	Type of event	Homoeo-SNP origin	Rec breakpoint positions	Nonsynonymous conversions (SNP position in ORF)
Cotton16_00047_02	AD1 A	Autapomorphy	–	–	–
	AD1 D	Autapomorphy	–	–	–
Cotton16_00080_04	AD1 A	–	AAAA	–	–
	AD1 D	Single recBP	DDDA	705	Q to L (689)
Cotton16_00255_03	AD1 A	Double recBP	aaaaaaaaAAAAaaaddddddaaaaaa	1217–1387	–
	AD1 D	3 or more recBP	aaaaaaaaddDDdaaaaddddddaaaaaa	562–919–1217–1387	–
Cotton16_01545_01	AD1 A	Paralogy	–	–	–
	AD1 D	Paralogy	–	–	–
Cotton16_01609_01	AD1 A	Single recBP	AaaAD	928	–
	AD1 D	–	DddDD-	–	–
Cotton16_01666_01	AD1 A	–	AAAAAA	–	–
	AD1 D	Single recBP	AAADDD	765	V to I (616)
Cotton16_01765_02	AD1 A	Autapomorphy	–	–	–
	AD1 D	Autapomorphy	–	–	–
Cotton16_01955_02	AD1 A	Single recBP	AaaaaadDDD	389	–
	AD1 D	Single recBP	DddddddDDA	705	–
Cotton16_02054_02	AD1 A	Autapomorphy	–	–	–
	AD1 D	Autapomorphy	–	–	–
Cotton16_07872_01	AD1 A	–	AAA	–	–
	AD1 D	Double recBP	DAD	1025–1122	–
Cotton16_10388_01	AD1 A	Single recBP	AAAdddd	363	–
	AD1 D	Single recBP	DDDaAAA	–	–
Cotton16_13240_01	AD1 A	Single recBP	DAAAA	1339	–
	AD1 D	–	DDDDD	–	–
Cotton16_13273_01	AD1 A	–	AAAAAA	–	–
	AD1 D	3 or more recBP	DADDDA	1676–2012–2158	–
Cotton16_14915_01	AD1 A	–	AAaaaaaaaaA	–	–
	AD1 D	Single recBP	AAdddddddddd	859	–
Cotton16_27501_01	AD1 A	3 or more recBP	DDDDDDADAAD	1100–1155–1172–1199	–
	AD1 D	3 or more recBP	DDDDDDADADD	1100–1155–1172–1178	R to P (536); Y to D (559)
Cotton16_34868_01	AD1 A	–	AA	–	–
	AD1 D	Single recBP	DA	1152	–
Cotton16_34881_01	AD1 A	Single recBP	AAAADddd	351	P to S (88)
	AD1 D	–	DDDDDDddd	–	–
Cotton28_21020	AD1 A	Autapomorphy	–	–	–
	AD1 D	Autapomorphy	–	–	–
Cotton28_21491	AD1 A	Double recBP	AAAADA	751–766	L to P (725)
	AD1 D	Double recBP	DDDDAD	751–766	P to L (725)
Cotton28_24190	AD1 A	Paralogy	–	–	–
	AD1 D	Paralogy	–	–	–

Homoeo-SNPs in capital letters refer to exonic sites, whereas those in lower case letters denote intronic sites.

Homoeo-SNP, homoeologous SNP origin; ORF, open reading frame; recBP, recombination breakpoint.

with the intronic sequences being part of the recombined regions, as opposed to the alternative explanation, whereby the EST evidence reflects post-transcriptional interactions (e.g. RNA editing). Second, the intronic sequences revealed a case in which an NRHR happened within an intron (Cotton16_00255_03; positions 1217–1387, Table 3).

Of the 14 regions in which homoeologous exchanges were confirmed, 10 regions displayed NRHR, six being D homoeologs that had experienced conversion of a region by the A homoeolog, and the four others exhibiting the reciprocal situation. Finally, four additional regions exhibited

more complex patterns, reflecting a history of more than two recombination breakpoints. Two of these regions involved NRHR patterns among the two copies detected: the Cotton16_00255_03 region had two copies, one displaying an intronic recombination pattern and the other displaying two recombinant regions, and the Cotton16_001955_02 region presented a suspected A-genome copy with an intronic recombination breakpoint (position 389), whereas the D-genome copy displayed an exonic recombination breakpoint (position 705). The two others (Cotton16_010388_01 and Cotton28_21491) involved

Table 4 Sequences cloned from six selected regions displaying a validated homoeologous recombination event in the four other AD-genome allopolyploids

Contig	Genome origin	Type of event	Homoeo-SNP origin pattern	Rec. breakpoint positions	Nonsynonymous conversions (position in ORF)
Cotton16_00255_03	AD3 D	Single recBP	dddddddddDDDDdaaaaaaaaaaaaaa	1088	–
	AD4 A	NA	–	–	–
	AD4 D	NA	–	–	–
Cotton16_01609_01	AD2 D	Single recBP	AddDD	427	A to V (347)
	AD3 D	Single recBP	AddDD	427	A to V (347)
Cotton16_01955_02	AD2 A	Single recBP	DdaaaaaAAA	239	–
	AD3 D	Missing copy	–	–	–
Cotton16_10388_01	AD2 A	Single recBP	DDDaaaa	363	G to D (272)
	AD5 A	NA	–	–	–
	AD5 D	NA	–	–	–
Cotton16_14915_01	AD3 D	Single recBP	AAAAadddddD	926	–
Cotton16_34881_01	AD5 A	Double recBP	AAAADDaA	351–438	–

Taxon designations are as follows: AD2, *Gossypium barbadense*; AD3, *G. tomentosum*; AD4, *G. mustelinum*; AD5, *G. darwinii*. Homoeo-SNPs in capital letters refer to exonic sites, whereas those in lower case letters denote intronic sites. Homoeo-SNP, homoeologous SNP origin; ORF, open reading frame; RecBP, recombination breakpoint.

the presence of reciprocal A-converted and D-converted copies that could originate from reciprocal double COs.

A logical first step in assessing the likelihood that homoeologous exchanges might have physiological relevance or functional consequences is to ask whether the conversions observed lead to amino acid alterations in the predicted translation products. In our dataset, such amino acid replacements were observed for six regions (nonsynonymous conversions in Table 3 for Cotton16_00080_04, Cotton16_01666_01, Cotton16_01765_02, Cotton16_27501_01, Cotton16_34881_01, Cotton28_21491).

Timing of homoeologous recombination following allopolyploid formation

In an effort to address the question of whether homoeologous recombination is a phenomenon restricted to the early stages of allopolyploid genome stabilization or, instead,

occurs more evenly during and after speciation, we assayed for the presence of putative homoeologous recombinants by amplifying, cloning and sequencing both homoeologs from the four remaining allopolyploid species. In addition, we also resequenced the homologous regions from both parental diploids, confirming the presence of only one allelic type for each region in each parental species, and confirming the homoeo-SNPs detected in the EST assembly. For each of these six regions, all A- and D-genome homoeologs were recovered, with the single exception of the AD3 D homoeolog of the Cotton16_01955_02 region.

For all six regions studied in this broader sampling of species, at least one other allopolyploid exhibited a homoeologous recombinant (Table 4). To place these in a phylogenetic context, we superimposed the placement of these putative events onto a cladogram (Fig. 2) that depicts organismal relationships (Wendel & Cronn, 2003), numbering the six contigs I–VI, as shown in Table 4.

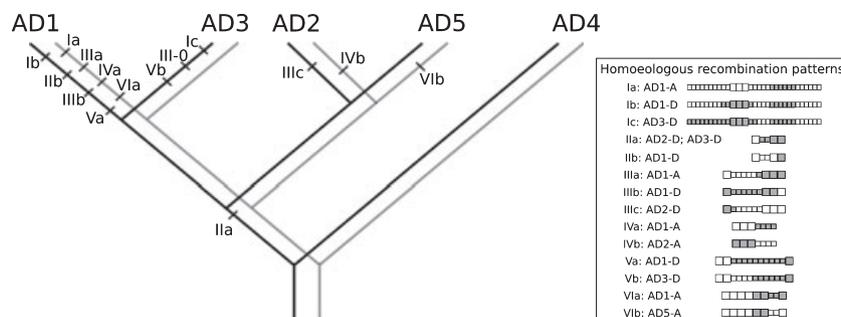


Fig. 2 Phylogenetic relationships among the five AD-genome polyploid species (A-genome homoeologs in grey; D-genome homoeologs in black) and homoeologous recombination events and their patterns detected in six regions and described in Table 4 (I, Cotton16_00255_03; II, Cotton16_01609_01; III, Cotton16_01955_02; IV, Cotton16_10388_01; V, Cotton16_14915_01; VI, Cotton16_34881_01). White and grey boxes denote A-genome and D-genome single nucleotide polymorphisms (SNPs), respectively, whereas the box size denotes location (large, exon; small, intron).

Perhaps the most striking aspect of this expanded analysis was that not a single homoeologous recombinant appeared to be shared among species. Instead, all recombinations between homoeologous regions corresponded to homoeologous SNP patterns that were different from those originally detected in the *G. hirsutum* genome. Eight recombinant copies were judged to be specific to *G. hirsutum*, whereas three were specific to *G. tomentosum* (AD3), its closest relative. For these regions, only two recombinants were specific to *G. barbadense* (AD2), one was restricted to *G. darwinii*, and not a single novel recombinant was detected in *G. mustelinum*. It is important to note that these data do not imply any overall difference in the prevalence of recombination between homoeologous loci in these species, as there is a strong ascertainment bias in the regions chosen (i.e. they were selected on the basis of evidence for homoeologous recombination in *G. hirsutum*). Yet, the data do underscore the fact that none of the regions selected for phylogenetic analysis diagnosed an ancient event. In fact, only a single copy was phylogenetically mapped as shared (IIa, Fig. 2). This event, shared by AD2 and AD3, may be phylogenetically complex, in that it occupies a position basal to their common ancestor. Hence, the AD1 recombinant copy may be derivative of this earlier recombination event.

Discussion

Polyploid speciation is a prominent mode of cladogenesis in flowering plants, and one that is now widely recognized to variously involve multiple non-Mendelian processes, including the loss of DNA (Ozkan *et al.*, 2001; Shaked *et al.*, 2001; Gaeta *et al.*, 2007; Grover *et al.*, 2008; Tate *et al.*, 2009), chromosomal rearrangements (Pires *et al.*, 2004; Pontes *et al.*, 2004; Udall *et al.*, 2005), repatterning of epigenetic marks (Madlung *et al.*, 2002; Liu & Wendel, 2003; Wang *et al.*, 2004; Rapp & Wendel, 2005; Salmon *et al.*, 2005; Lukens *et al.*, 2006; Chen, 2007) and biased gene expression (Adams *et al.*, 2003, 2004; Albertin *et al.*, 2006; Hegarty *et al.*, 2006; Tate *et al.*, 2006; Wang *et al.*, 2006; Gaeta *et al.*, 2007; Flagel *et al.*, 2008; Ha *et al.*, 2009; Rapp *et al.*, 2009). One aspect of polyploidy that has received relatively little attention is the possibility of nonreciprocal homoeologous recombination throughout the genome, with only rDNA loci being commonly surveyed in this respect (Wendel *et al.*, 1995; Kovarik *et al.*, 2008).

Our analysis of homoeologous recombinants in the AD-genome polyploid cottons allowed us to estimate the proportion of both homoeologous reciprocal and nonreciprocal recombinants in the cotton transcriptome. We took into consideration different artifacts that could affect this inference from EST assemblies: autapomorphies, EST assembly artifacts and post-transcriptional recombinations, by validating recombination events using *de novo* genomic DNA sequencing and intronic information. One other limitation

of the use of EST assemblies to detect gene conversion events relies in the fact that these represent the expressed part of the genome (thus missing all nonexpressed conversions or recombinants).

The recombinant homoeologous copies detected correspond to the formation of chimeric copies that represent, in some cases, validated nonsynonymous recombinants when compared with the parental copies. This result highlights the putative source of novelty of recombination of homoeologous copies on duplicated genes (contrary to the concerted evolution observed in specific gene families displaying gene conversion). In addition, the size of NRHR events, ranging from a few nucleotides to 700 nucleotides in length (for copies displaying two recombination breakpoints), is consistent with a previous report in rice (*Oryza sativa indica*), in which gene-converted regions had an average length of 130 nucleotides, ranging in size from 4 bp to more than 1 kb (Xu *et al.*, 2008). This latter observation, superimposed on the fact that reciprocal homoeologous recombination has not been observed in cotton and, indeed, is not evident in their genetic maps (Endrizzi *et al.*, 1985; Rong *et al.*, 2004), suggests that most of the NRHRs detected are a result of gene conversion events (NCOs) rather than COs between homoeologous chromatids. Wang *et al.* (2009) have also recently reported a comparison of rice and *Sorghum* genomes using phylogenetic analysis, which revealed gene conversion among paralogs within each species, leading to an accelerated divergence for each genome. The estimates of gene conversion among duplicated genes in rice and sorghum (5.5% and 4.1%, respectively) and in mammals (1–3%; McGrath *et al.*, 2009) appear to be in accordance with that reported here, although the previous estimates involve paralogous copies in a diploid, whereas our study focused on homoeologs in a polyploid.

The six regions displaying homoeologous exchanges in *G. hirsutum* were also detected in at least one of the other AD-genome polyploid species. It also appeared that the number of recombinant copies in the same region decreased in correspondence with the phylogenetic distance from *G. hirsutum*. This could reflect the gradual (as opposed to episodic) accumulation of homoeologous exchanges when coupled with speciation among these five polyploid species over the last 1.5 Myr. Homoeologous pairing during meiosis is suspected to be avoided in the first generations following polyploid formation (Ramsey & Schemske, 2002), because it can lead to infertility as a result of unbalanced gamete formation following meiosis. It is also clear, however, that some recombinations of homoeologous loci may in fact be part of polyploid stabilization in some lineages (Udall *et al.*, 2005). Our finding of a gradual accumulation of homoeologous exchanges is not surprising, as homoeologous pairing is likely to be rarely tolerated. In addition, the parental genomes of the allopolyploid cottons differ significantly in size (by a factor of approximately two;

Hendrix & Stewart, 2005), which has been shown to be mainly a result of transposable element insertions in the A genome (Hawkins *et al.*, 2006; Grover *et al.*, 2007). This large size discrepancy probably contributes to the difficulty in homoeologous pairing, and could further explain the late arrival and gradual accumulation of homoeologous exchanges during the evolutionary history of allopolyploid cotton. Although it is true that homoeologous pairing is rarely observed among modern stabilized cotton allopolyploids, classic cytogenetic experiments have demonstrated that it is more frequent in *Gossypium* neopolyploids (reviewed in Endrizzi *et al.*, 1985). Thus, one might expect that temporal analysis of recombination of homoeologous loci in *Gossypium* would have led to the discovery of primarily basal, shared events. Yet, our data do not show this effect. Instead, it appears that homoeologous recombinations have arisen sporadically, and have continued even after allopolyploid speciation. This observation is in agreement with gene conversion events rather than reciprocal homoeologous recombination, as NCOs are resolved before chiasmatic associations of homoeologous chromosomes during meiosis (Allers & Lichten, 2001).

Clearly, more work is needed on the pace and timing of such homoeologous exchanges over a longer timescale during and after polyploid formation. The most informative studies are likely to involve a combination of synthetic hybrids and allopolyploids, as well as natural allopolyploids, for which a well-understood phylogenetic framework exists. These types of studies may ultimately permit a model to be developed that would tie together homoeologous pairing over the generations, NRHR leading to gene conversion and the process of polyploid stabilization. In addition to these experiments, additional research is necessary to evaluate the provocative speculation that low sequence divergence between duplicated copies after polyploidization and conservation of duplicate gene expression may be causally connected to homoeologous gene conversion (Chapman *et al.*, 2006). Finally, and in spite of the generally homogenizing consequences of gene conversion, it remains to be demonstrated that the process leads to evolutionary novelty, either by generating chimeric gene copies, or by fixing across two homoeologs favored amino acids that previously were restricted to only one of the two gene copies.

In this study, we found NRHRs in a small percentage of the cotton genome. It is not clear whether the function of these genes is related to the frequency of gene conversion. Continued sequencing of ESTs and the cotton genome will improve our ability to relate the significance of gene conversion to genome evolution and adaptation.

Acknowledgements

We gratefully acknowledge the National Research Initiative of the USDA Cooperative State Research, Education and

Extension Service (2005-35301-15700 to J.A.U and J.F.W.), the National Science Foundation Plant Genome Research Program (0638418 to J.F.W.) and the Department of Ecology, Evolution, and Organismal Biology for a Graduate Research Award to LF. We thank K. Grupp and A. Balu for technical assistance.

References

- Adams KL, Cronn R, Percifield R, Wendel JF. 2003. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proceedings of the National Academy of Sciences, USA* **100**: 4649–4654.
- Adams KL, Percifield R, Wendel JF. 2004. Organ-specific silencing of duplicated genes in a newly synthesized cotton allotetraploid. *Genetics* **168**: 2217–2226.
- Albertin W, Balliau T, Brabant P, Chevre AM, Eber F, Malosse C, Thiellement H. 2006. Numerous and rapid nonstochastic modifications of gene products in newly synthesized *Brassica napus* allotetraploids. *Genetics* **173**: 1101–1113.
- Allers T, Lichten M. 2001. Differential timing and control of non-crossover and crossover recombination during meiosis. *Cell* **106**: 47–57.
- Chapman BA, Bowers JE, Feltus FA, Paterson AH. 2006. Buffering of crucial functions by paleologous duplicated genes may contribute cyclicity to angiosperm genome duplication. *Proceedings of the National Academy of Sciences, USA* **103**: 2730–2735.
- Chen ZJ. 2007. Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annual Review of Plant Biology* **58**: 377–406.
- Chen JM, Cooper DN, Chuzhanova N, Ferec C, Patrinos GP. 2007. Gene conversion: mechanisms, evolution and human disease. *Nature Reviews Genetics* **8**: 762–775.
- Conant GC, Wolfe KH. 2008. Turning a hobby into a job: how duplicated genes find new functions. *Nature Reviews Genetics* **9**: 938–950.
- Cronn RC, Small RL, Wendel JF. 1999. Duplicated genes evolve independently after polyploid formation in cotton. *Proceedings of the National Academy of Sciences, USA* **96**: 14406–14411.
- Cronn R, Cedroni M, Haselkorn T, Grover C, Wendel JF. 2002. PCR-mediated recombination in amplification products derived from polyploid cotton. *Theoretical and Applied Genetics* **104**: 482–489.
- Doyle JJ. 1991. DNA protocols for plants. In: Hewitt GM, Johnston A, eds. *Molecular Techniques in Taxonomy*. Berlin: Springer-Verlag, 283–293.
- 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.
- Endrizzi JE, Turcotte EL, Kohel RJ. 1985. Genetics, cytology and evolution of *Gossypium*. *Advances in Genetics* **23**: 271–375.
- Flagel L, Wendel JF. 2009. Gene duplication and evolutionary novelty in plants. *New Phytologist* **183**: 557–563.
- Flagel L, Udall J, Nettleton D, Wendel J. 2008. Duplicate gene expression in allopolyploid *Gossypium* reveals two temporally distinct phases of expression evolution. *BMC Biology* **6**: 16.
- 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. 2007. Microcolinearity and genome evolution in the adha region of diploid and polyploid cotton (*Gossypium*). *Plant Journal* **50**: 995–1006.
- Grover CE, Yu Y, Wing RA, Paterson AH, Wendel JF. 2008. A phylogenetic analysis of indel dynamics in the cotton genus. *Molecular Biology and Evolution* **25**: 1415–1428.

- Ha M, Kim ED, Chen ZJ. 2009. Duplicate genes increase expression diversity in closely related species and allopolyploids. *Proceedings of the National Academy of Sciences, USA* 106: 2295–2300.
- Hawkins JS, Kim H, Nason JD, Wing RA, Wendel JF. 2006. Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. *Genome Research* 16: 1252–1261.
- Hegarty MJ, Barker GL, Wilson ID, Abbott RJ, Edwards KJ, Hiscock SJ. 2006. Transcriptome shock after interspecific hybridization in senescio is ameliorated by genome duplication. *Current Biology* 16: 1652–1659.
- Hendrix B, Stewart JM. 2005. Estimation of the nuclear DNA content of *Gossypium* species. *Annals of Botany* 95: 789–797.
- Hollingsworth NM, Brill SJ. 2004. The mus81 solution to resolution: generating meiotic crossovers without holliday junctions. *Genes & Development* 18: 117–125.
- Judo MSB, Wedel AB, Wilson C. 1998. Stimulation and suppression of PCR-mediated recombination. *Nucleic Acids Research* 26: 1819–1825.
- Kovarik A, Matyasek R, Lim KY, Skalicka K, Koukalova B, Knapp S, Chase M, Leitch AR. 2004. Concerted evolution of 18-5.8-26s rDNA repeats in *Nicotiana* allotetraploids. *Biological Journal of the Linnean Society* 82: 615–625.
- Kovarik A, Pires JC, Leitch AR, Lim KY, Sherwood AM, Matyasek R, Rocca J, Soltis DE, Soltis PS. 2005. Rapid concerted evolution of nuclear ribosomal DNA in two *Tragopogon* allopolyploids of recent and recurrent origin. *Genetics* 169: 931–944.
- Kovarik A, Dadejova M, Lim YK, Chase MW, Clarkson JJ, Knapp S, Leitch AR. 2008. Evolution of rDNA in *Nicotiana* allopolyploids: a potential link between rDNA homogenization and epigenetics. *Annals of Botany* 101: 815–823.
- Krapovickas A, Seijo G. 2008. *Gossypium ekmanianum* (malvaceae), algodon silvestre de la República Dominicana. *Bonplandia* 17: 55–63.
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R *et al.* 2007. Clustal w and clustal x version 2.0. *Bioinformatics* 23: 2947–2948.
- Liu B, Wendel JF. 2003. Epigenetic phenomena and the evolution of plant allopolyploids. *Molecular Phylogenetics and Evolution* 29: 365–379.
- Lukens LN, Pires JC, Leon E, Vogelzang R, Oslach L, Osborn T. 2006. Patterns of sequence loss and cytosine methylation within a population of newly resynthesized *Brassica napus* allopolyploids. *Plant Physiology* 140: 336–348.
- Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290: 1151–1155.
- Madlung A, Masuelli RW, Watson B, Reynolds SH, Davison J, Comai L. 2002. Remodeling of DNA methylation and phenotypic and transcriptional changes in synthetic *Arabidopsis* allotetraploids. *Plant Physiology* 129: 733–746.
- Martinez-Perez E, Colaiacovo MP. 2009. Distribution of meiotic recombination events: talking to your neighbors. *Current Opinion in Genetics & Development* 19: 105–112.
- de Massy B. 2003. Distribution of meiotic recombination sites. *Trends in Genetics* 19: 514–522.
- McGrath CL, Casola C, Hahn MW. 2009. Minimal effect of ectopic gene conversion among recent duplicates in four mammalian genomes. *Genetics* 182: 615–622.
- Mezard C, Vignard J, Drouaud J, Mercier R. 2007. The road to crossovers: plants have their say. *Trends in Genetics* 23: 91–99.
- Nicolas SD, Leflon M, Liu Z, Eber F, Chelysheva L, Coriton O, Chevre AM, Jenczewski E. 2008. Chromosome ‘speed dating’ during meiosis of polyploid *Brassica* hybrids and haploids. *Cytogenetic and Genome Research* 120: 331–338.
- Ohno S. 1970. *Evolution by gene duplication*. New York, NY, USA: Springer-Verlag.
- Ozkan H, Levy AA, Feldman M. 2001. Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. *Plant Cell* 13: 1735–1747.
- Paques F, Haber JE. 1999. Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiology and Molecular Biology Reviews* 63: 349–404.
- Parkin IAP, Sharpe AG, Keith DJ, Lydiat DJ. 1995. Identification of the A and C genomes of amphidiploid *Brassica napus* (oilseed rape). *Genome* 38: 1122–1131.
- Pires JC, Zhao JW, 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.
- Pontes O, Neves N, Silva M, Lewis MS, Madlung A, Comai L, Viegas W, Pikaard CS. 2004. Chromosomal locus rearrangements are a rapid response to formation of the allotetraploid *Arabidopsis suecica* genome. *Proceedings of the National Academy of Sciences, USA* 101: 18240–18245.
- Ramsey J, Schemske DW. 2002. Neopolyploidy in flowering plants. *Annual Review of Ecology and Systematics* 33: 589–639.
- 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.
- Rong JK, Abbey C, Bowers JE, Brubaker CL, Chang C, Chee PW, Delmonte TA, Ding XL, Garza JJ, Marler BS *et al.* 2004. A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). *Genetics* 166: 389–417.
- Rozen S, Skaletsky H. 2000. “Primer3 on the www for general users and for biologist programmers”. In: Krawetz S, Misener S, eds. *Bioinformatics methods and protocols*. Totowa, NJ, USA: Humana Press, 365–386.
- Salmon A, Ainouche ML, Wendel JF. 2005. Genetic and epigenetic consequences of recent hybridization and polyploidy in *Spartina* (Poaceae). *Molecular Ecology* 14: 1163–1175.
- Seelanan T, Schnabel A, Wendel JF. 1997. Congruence and consensus in the cotton tribe (Malvaceae). *Systematic Botany* 22: 259–290.
- Senchina DS, Alvarez I, Cronn RC, Liu B, Rong JK, 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, Wendel JF. 1999. The mitochondrial genome of allotetraploid cotton (*Gossypium* L.). *Journal of Heredity* 90: 251–253.
- 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.
- Stephens SG. 1951. Possible significance of duplication in evolution. *Advances in Genetics Incorporating Molecular Genetic Medicine* 4: 247–265.
- Szostak JW, Orrweaver TL, Rothstein RJ, Stahl FW. 1983. The double-strand-break repair model for recombination. *Cell* 33: 25–35.
- Tate JA, Ni ZF, 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: 10.
- Teshima KM, Innan H. 2004. The effect of gene conversion on the divergence between duplicated genes. *Genetics* 166: 1553–1560.
- Udall JA, Quijada PA, Osborn TC. 2005. Detection of chromosomal rearrangements derived from homeologous recombination in four mapping populations of *Brassica napus* L. *Genetics* 169: 967–979.

- Udall JA, Swanson JM, Haller K, Rapp RA, Sparks ME, Hatfield J, Yu YS, Wu YR, 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.
- Wang JL, Tian L, Madlung A, Lee HS, Chen M, Lee JJ, Watson B, Kagochi T, Comai L, Chen ZJ. 2004. Stochastic and epigenetic changes of gene expression in *Arabidopsis* polyploids. *Genetics* 167: 1961–1973.
- Wang JL, Tian L, Lee HS, Wei NE, Jiang HM, Watson B, Madlung A, Osborn TC, Doerge RW, Comai L *et al.* 2006. Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. *Genetics* 172: 507–517.
- Wang XY, Tang HB, Bowers JE, Paterson AH. 2009. Comparative inference of illegitimate recombination between rice and sorghum duplicated genes produced by polyploidization. *Genome Research* 19: 1026–1032.
- Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview version 2 – a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25: 1189–1191.
- Wendel JF, Cronn RC. 2003. Polyploidy and the evolutionary history of cotton. *Advances in Agronomy* 78: 139–186.
- Wendel JF, Schnabel A, Seelanan T. 1995. An unusual ribosomal DNA-sequence from *Gossypium gossypoides* reveals ancient, cryptic, intergenomic introgression. *Molecular Phylogenetics and Evolution* 4: 298–313.
- Xu SQ, Clark T, Zheng HK, Vang S, Li RQ, Wong GKS, Wang J, Zheng XG. 2008. Gene conversion in the rice genome. *BMC Genomics* 9: 93.

Supporting Information

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

Table S1 List of primers designed to amplify selected contigs displaying putative NRHR

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.



About *New Phytologist*

- *New Phytologist* is owned by a non-profit-making **charitable trust** dedicated to the promotion of plant science, facilitating projects from symposia to open access for our Tansley reviews. Complete information is available at www.newphytologist.org.
- Regular papers, Letters, Research reviews, Rapid reports and both Modelling/Theory and Methods papers are encouraged. We are committed to rapid processing, from online submission through to publication 'as-ready' via *Early View* – our average submission to decision time is just 29 days. Online-only colour is **free**, and essential print colour costs will be met if necessary. We also provide 25 offprints as well as a PDF for each article.
- For online summaries and ToC alerts, go to the website and click on 'Journal online'. You can take out a **personal subscription** to the journal for a fraction of the institutional price. Rates start at £151 in Europe/\$279 in the USA & Canada for the online edition (click on 'Subscribe' at the website).
- If you have any questions, do get in touch with Central Office (newphytol@lancaster.ac.uk; tel +44 1524 594691) or, for a local contact in North America, the US Office (newphytol@ornl.gov; tel +1 865 576 5261).