ORIGINAL ARTICLE

Gene loss and silencing in *Tragopogon miscellus* (Asteraceae): comparison of natural and synthetic allotetraploids

RJA Buggs^{1,5}, AN Doust^{1,5,6}, JA Tate^{1,7}, J Koh¹, K Soltis¹, FA Feltus^{2,8}, AH Paterson², PS Soltis^{3,4} and DE Soltis^{1,4}

¹Department of Biology, University of Florida, Gainesville, FL, USA; ²Plant Genome Mapping Laboratory, University of Georgia, Athens, GA, USA; ³Florida Museum of Natural History, University of Florida, Gainesville, FL, USA and ⁴Genetics Institute, University of Florida, Gainesville, FL, USA

Whole-genome duplication (polyploidisation) is a widespread mechanism of speciation in plants. Over time, polyploid genomes tend towards a more diploid-like state, through downsizing and loss of duplicated genes (homoeologues), but relatively little is known about the timing of gene loss during polyploid formation and stabilisation. Several studies have also shown gene transcription to be affected by polyploidisation. Here, we examine patterns of gene loss in 10 sets of homoeologues in five natural populations of the allotetraploid *Tragopogon miscellus* that arose within the past 80 years following independent whole-genome duplication events. We also examine 44 first-generation synthetic allopolyploids of the same species. No cases of homoeolo-

gue loss arose in the first allopolyploid generation, but after 80 years, 1.6% of homoeologues were lost in natural populations. For seven homoeologue sets we also examined transcription, finding that 3.4% of retained homoeologues had been silenced in the natural populations, but none in the synthetic plants. The homoeologue losses and silencing events found were not fixed within natural populations and did not form a predictable pattern among populations. We therefore show haphazard loss and silencing of homoeologues, occurring within decades of polyploid formation in *T. miscellus*, but not in the initial generation.

Heredity (2009) **103**, 73–81; doi:10.1038/hdy.2009.24; published online 11 March 2009

Keywords: polyploidy; whole-genome duplication; *Tragopogon miscellus*; homoeologue evolution; Ohnologue

Introduction

Polyploidisation, the duplication of whole genomes, has long been known as a major mechanism of abrupt speciation in plants, and current research shows it to have important consequences for subsequent genome and transcriptome evolution (Soltis *et al.*, 2004b; Comai, 2005; Hegarty and Hiscock, 2008; Leitch and Leitch, 2008). It can be difficult to untangle the effects of polyploidisation *per se* from the effects of hybridisation, which often accompanies it, and/or from the effects of evolutionary changes in subsequent generations. To do this, examination is needed of every stage from progenitor diploids through young polyploids to ancient

polyploids, but no single natural system currently presents such a series. Instead, we must use different species of natural polyploids of different ages to investigate the temporal dynamics of polyploid evolution, often using artificial crosses to investigate the effects of hybridisation, and synthetic polyploids to investigate the instantaneous effects of genome duplication.

Such studies have shown that, at the level of the genome, polyploids tend towards a more diploid-like state, through downsizing and loss of duplicated genes (homoeologues) (Wolfe, 2001; Leitch and Bennett, 2004; Chen *et al.*, 2007; Grover *et al.*, 2007). Fragments of DNA seem to have been lost through homoeologous recombination in synthetic allopolyploids (Liu *et al.*, 1998; Ozkan and Feldman, 2001; Shaked *et al.*, 2001; Kashkush *et al.*, 2002; Ozkan *et al.*, 2002; Gaeta *et al.*, 2007). ‘Genomic shock’ caused by whole-genome duplication may lead to gene silencing (McClintock, 1984; Comai *et al.*, 2003), and changes to the transcriptome have been found in synthetic allopolyploids (Wang *et al.*, 2004; Hegarty and Hiscock, 2008). However, we still know relatively little about natural genomic and transcriptomic changes during the early generations after polyploid formation, mainly because of the scarcity of young natural polyploid species for which genetic resources are available (Buggs, 2008).

The polyploid plant species *Tragopogon miscellus* provides an excellent evolutionary model (Soltis *et al.*,

Correspondence: Dr RJA Buggs, Department of Biology, University of Florida, Gainesville, FL 32611, USA.
E-mail: buggs@ufl.edu

⁵These authors contributed equally to this work.

⁶Current address: Botany Department, Oklahoma State University, 104 Life Sciences East, Stillwater, OK 74078, USA.

⁷Current address: Massey University, Institute of Molecular Biosciences, Private Bag 11222, Palmerston North, New Zealand.

⁸Current address: Clemson University, Department of Genetics and Biochemistry, Biosystems Research Complex, 51 New Cherry Street, Clemson, SC 29634, USA.

Received 29 October 2008; accepted 3 February 2009; published online 11 March 2009

2004b) in which to investigate evolution in the early generation's post-polyploidisation. It arose naturally <80 years ago through allopolyploidisation, with the diploid species *T. dubius* and *T. pratensis* as parents (Ownbey, 1950; Soltis *et al.*, 2004a). These parents are phylogenetically divergent (Mavrodiev *et al.*, 2005; Buggs *et al.*, 2008); therefore, homoeologues in *T. miscellus* can be distinguished by sequence differences (for example, Tate *et al.*, 2006). The species formed repeatedly in different localities from separate populations of the diploid progenitors (reviewed in Soltis *et al.*, 1995, 2004a), giving replicated independent natural allopolyploid lines. In addition, we now have multiple, reciprocal synthetic allopolyploid lines of the species (Tate *et al.*, in press). These resources allow us to examine the instantaneous effects of allopolyploidisation, as well as its effects over 40 generations (80 years) in this biennial species.

Here, we use the *T. miscellus* system to examine the rapid evolution of 10 sets of homoeologous genes in multiple individuals from five independent natural lineages and 44 first-generation synthetic allopolyploids. Distinguishing between homoeologues using single-nucleotide polymorphisms (SNPs), we ask whether loss or silencing of these genes has occurred immediately on polyploidisation (that is, within the first generation), more gradually but within 40 generations, or not at all within so short a time frame? We also examine the possibility that there are concerted mechanisms for gene loss, by determining whether there are repeated patterns of loss or silencing among independent origins.

Materials and methods

Plant materials

Seeds from five natural populations of *T. miscellus* (Table 1), together with the seed of the diploid parent species, *T. dubius* and *T. pratensis*, were germinated and grown under controlled conditions in a greenhouse at the University of Florida (Gainesville, FL, USA). The five natural populations of *T. miscellus* all formed independently and included lineages of reciprocal origin: the short-liguled form (with *T. pratensis* as the maternal parent) from Moscow, ID, and Garfield, Oakesdale and Spangle, WA, USA, and the long-liguled form (with

T. dubius as the maternal parent) from Pullman, WA, USA (Soltis and Soltis collection numbers for all populations are given in Table 1). Samples of *T. dubius* and *T. pratensis* were obtained from the same locations (except Pullman, where *T. pratensis* is not currently found (Novak *et al.*, 1991)) and analysed with their progeny to ensure that the parents of the allopolyploid lineages were not heterozygous for the markers surveyed. Field-collected seeds were used from Oakesdale, Spangle and Garfield populations; seeds used from Moscow and Pullman populations were derived from plants grown in the greenhouse (at Washington State University, Pullman, WA, USA) from field-collected seeds and allowed to self-fertilise for one generation.

In addition, synthetic allopolyploids, formed from reciprocal hybridisation between *T. dubius* and *T. pratensis* followed by colchicine treatment (Tate *et al.*, in press), were examined. Flowering was induced by cold treating the raw synthetic allopolyploids (S_0 generation) for 3 months in a growth chamber at 10 °C under short-day-length conditions (8 h of light). After 3 months, the plants were placed back in the greenhouse and allowed to flower and self-fertilise. Capitula were enclosed in glassine envelopes (Paper Mart, Los Angeles, CA, USA) before the heads opening and maintained thus until mature achenes formed. Seeds from these initial synthetic allopolyploids (S_1 generation) were germinated and grown in the same manner as in the case of the earlier generation. Leaf tissue for this study was sampled from 44 S_1 plants, representing progeny from six initial crosses (see Supplementary Material, Table S1).

DNA and RNA extraction

Leaf tissue for the natural diploid and allopolyploid populations and synthetic allopolyploids was collected from seedlings 4 weeks after germination and flash frozen in liquid nitrogen (Tate *et al.*, 2006). DNA was extracted from leaf tissue using a modified cetyl trimethylammonium bromide protocol (Doyle and Doyle, 1987), after homogenising tissue in a bead mill (BioSpec, Bartlesville, OK, USA). RNA was extracted using the RNeasy kit from Qiagen (Valencia, CA, USA). RNA was DNase-d to eliminate residual genomic DNA using the DNA-free DNase kit from Ambion (Austin, TX, USA). First-strand cDNA synthesis was carried out on 500 ng RNA using superscript II reverse transcriptase (Invitrogen, Carlsbad, CA, USA) and random hexamer primers.

Homoeologue identification

Arabidopsis thaliana genes (Paterson *et al.*, 2006) were tblastn-aligned to 2081 *Tragopogon*, 227 016 *Lactuca* and 146 701 *Helianthus* expressed sequence tags (ESTs), with a minimum criterion of $E < 0.001$. After filtering for $\leq 50\%$ identity, 807 *Lactuca*, 0 *Tragopogon* and 262 *Helianthus* ESTs matched at the translated protein level. The best *Lactuca* EST was then determined for 78 *A. thaliana* genes and blastn-aligned to *A. thaliana* genomic DNA (minimum $E < 1 \times 10^{-10}$). From this search, 22 *A. thaliana* loci were aligned to these ESTs, each of which spanned one to three introns (based on the *A. thaliana* primary sequence).

Primer design and PCR

Primers were designed both directly from the *Arabidopsis* genomic sequence using primer 3 (Rozen and Skaletsky,

Table 1 Natural populations analysed

Population	Species	Collection number	Number of individuals analysed		Ligule length
			gDNA	cDNA	
Moscow	<i>T. pratensis</i>	2608	10	10	—
	<i>T. miscellus</i>	2604	18	10	Short
	<i>T. dubius</i>	2631	5	5	—
Pullman	<i>T. pratensis</i>	2608	—	—	—
	<i>T. miscellus</i>	2605	15	10	Long
	<i>T. dubius</i>	2613	10	9	—
Garfield	<i>T. pratensis</i>	2689	3	5	—
	<i>T. miscellus</i>	2688	8	5	Short
	<i>T. dubius</i>	2687	3	5	—
Oakesdale	<i>T. pratensis</i>	2672	8	5	—
	<i>T. miscellus</i>	2671	8	5	Short
	<i>T. dubius</i>	2670	7	5	—
Spangle	<i>T. pratensis</i>	2692	6	5	—
	<i>T. miscellus</i>	2693	8	5	Short
	<i>T. dubius</i>	2616	6	5	—

Table 2 Homoeologous loci examined in *T. miscellus* with genomic and cDNA CAPS analysis

Homoeologue pair	Putative function	Primers	Genbank accession numbers	Diagnostic restriction enzyme	Genomic fragment sizes	cDNA fragment sizes
S2	RNA binding	F: TYACTCAYGTYTCAMGAGGATTTG R: AGTAATAACCAGATGCTTCATCCC	<i>T. dubius</i> S2A FJ708502 <i>T. pratensis</i> S2B FJ708503 Region amplified is from bp 18 to 256 (gDNA) and 18 to 173 (cDNA) of FJ708502 and FJ708503	<i>Dra</i> I	T.d. 362/160 T.p. 527	
S3	NADP/FAD oxidoreductase	Fp: GAGGATTGTCCTTTGTACATTTTCATT Fd: CCTAGAGGCAACAAATGGGACAACCGTA Rp: ACACGTAAGATTGACCGTTTTTCACA Rd: AGTTACCTGTTGGGAAAATGTTGCAGC F: AGGCAAAGCACCTTCAAAGA R: CTTCCTACTGCCAGCTTTTTC	<i>T. dubius</i> S3A FJ708504 <i>T. pratensis</i> S3B FJ708505	<i>Mbo</i> II	T.d. 150 T.p. 187	T.d. 150 T.p. 106
S4	Biotin synthase	F: GGTCCGAGAACTGATTGGAA R: ATTGGGGGCAATATGAACAA	<i>T. dubius</i> S4A FJ708506 <i>T. pratensis</i> S4B FJ708507	<i>Dde</i> I	T.d. 593 T.p. 270/299	
S5	c-Myc protein	F: TCTTCAAGAYAGYCTCAMYAAACATCC R: TCYTTYCCWTCWTCCATTGTT	<i>T. dubius</i> S5A FJ708508 <i>T. pratensis</i> S5B FJ708509	<i>Alu</i> I (gDNA) <i>Mbo</i> II (cDNA)	T.d. 629/315 T.p. 1000	T.d. 217 T.p. 81/138
S7	Nucleic acid binding	F: WTGGGAAGAAGGMGAAGG R: TTATCRAACTGWGTWGTATCAAATGCC	<i>T. dubius</i> S7A FJ708510 <i>T. pratensis</i> S7B FJ708511	<i>Mbo</i> II	T.d. 719 T.p. 267/499	
S8	Acetyl transferase	F: TGGGYCACAGTGTAGATAAG R: AGCAACWGCYTCTTCAACGT	<i>T. dubius</i> S8A FJ708512 <i>T. pratensis</i> S8B FJ708513	<i>Fok</i> I	T.d. 229 T.p. 111/118	T.d. 146 T.p. 31/115
S13	Prenyltransferase	F: ARCCWGTMACTTGGCCTCC R: ATKGCTCCMGAWGGAATTGG	<i>T. dubius</i> 132A FJ708514 <i>T. pratensis</i> S13B FJ708515	<i>Nde</i> I	T.d. 287/99 T.p. 400	
S18	Porphyrin-oxidoreductase	F: TYTGYCGCAAACGCTGTC R: TCWCTCTGAAAWGCTTGAACCTCC	<i>T. dubius</i> S18A FJ708516 <i>T. pratensis</i> S18B FJ708517 Region amplified is from bp 18–256 (gDNA) of FJ708516 and FJ708517	<i>Ple</i> I	T.d. 209/167 T.p. 375	T.d. 209/167 T.p. 375
S22	SET domain containing protein	Fp: TGTCATTATTGTGACTTNCCAATACTTGC Fd: TCTCAAATTACATCCACATCCTTATTAGC Rp: AATGTGTTTGTGTGCTTCAATTGCA Rd: AACGTATCAAGAACCTTAGCGACGAG F: GKGATGSYGATGGWTATG R: GTCAAGTGTAASAGCYAAATTCTG	<i>T. dubius</i> S22A FJ708518 <i>T. pratensis</i> S22B FJ708519	<i>Bgl</i> II	T.d. 1034 T.p. 666/349	T.d. 642 T.p. 437/205
D1	LRR protein kinase	F: ATTGAAATTGTATATGCAGGGAAGGCCG R: ATCATCTCCCAACAATTGGACAATCCA	<i>T. dubius</i> D1A FJ708520 <i>T. pratensis</i> D1B FJ708521	<i>Hyp</i> CH4IV	T.d. 210 T.p. 83/127	T.d. 210 T.p. 83/127

Primer abbreviations: F, forward CAPS primer; Fd, forward *T. dubius* homoeologue-specific primer; Fp, forward *T. pratensis* homoeologue-specific primer; R, reverse CAPS primer; Rd, reverse *T. dubius* homoeologue-specific primer; Rp, reverse *T. pratensis* homoeologue-specific primer.
Putative gene functions are based on known functions of *Arabidopsis* orthologues. Fragment lengths include primer sequence.

1999) and from an alignment of *Arabidopsis* genomic and cDNA and *Lactuca* ESTs using Primacade (Gadberry *et al.*, 2005) (Table 2). Primers were positioned in exons and separated by at least one intron, so that genomic and cDNA amplicons would be of different sizes and distinguishable on agarose gels. We were able to successfully use these primers to amplify 10 genes from *Tragopogon* genomic DNA and nine genes from cDNA. In the case of gene S13, for which primers did not work with cDNA, several additional sets of primers were designed; yet, they failed to amplify in either polyploids or diploids, perhaps because the genes were not expressed in the rosette leaf tissue at the time of collection.

Genomic and cDNA fragments were amplified in 25 µl volume with 50–100 ng template, 20% final volume Promega 5× sequencing buffer (Promega, Madison, WI, USA), 10% final volume 5M betaine, 1 mM MgCl₂, 0.4 mM dNTPs, 0.2 µM of each primer and 0.4 units of Promega *Taq* polymerase. Most of the primers were designed with annealing temperatures close to 60 °C; a representative set of thermocycling conditions is as follows: 94 °C for 2 min, followed by two cycles of 94 °C for 30 s, 60 °C for 30 s and 72 °C for 1 min, followed by three further sets of two cycles in which the annealing temperature was dropped to 58, 56 and 54 °C for each set, followed by 27 cycles with an annealing temperature of 52 °C. Gene fragments were amplified from the two diploid parents, *T. dubius* and *T. pratensis*, and run out on agarose gels. The PCR product was excised from the agarose gel, cleaned using the Wizard SV gel clean-up kit (Promega) and cloned using the Topo-TA cloning kit (Invitrogen). Two to four clones were sequenced using M13 forward and reverse primers on an ABI 3730 capillary sequencer, and fragments were assembled into contigs using Sequencher version 4.8 (Gene Codes, Ann Arbor, MI, USA). Sequences of the genes successfully amplified from *T. dubius* and *T. pratensis* genomic DNA were submitted to GenBank (accession numbers FJ708502–FJ708521).

Sequence alignment, SNP detection and restriction site detection

Sequences were aligned using MacClade (Maddison and Maddison, 2004) and the SNPs identified. Each SNP was investigated, using DNA Strider 1.1 (Douglas, 1995), to determine whether it was part of a restriction site that would differentiate the two diploid parental species. In most cases, a restriction site difference was identified in at least one exon for each gene, and these were then used

to construct CAPS markers (Konieczny and Ausubel, 1993; Tate *et al.*, 2006). In an allopolyploid, there are two copies of each gene, with each homoeologue possessing two alleles (if an allopolyploid has been formed by genome duplication of an F₁ hybrid, the two alleles of each homoeologue will be identical). Both alleles of a given homoeologue must be lost for the loss to be visualised by CAPS analysis. Therefore, there may be some gene losses that went undetected in this study (for example, one of two copies of a gene).

In three cases, there was insufficient SNP variation between parental alleles in the exon sequences for selection of diagnostic restriction enzymes. In one of these cases, an exon SNP was exploited by designing primers that would amplify only one or the other of the SNP alleles (Ye *et al.*, 2001). In the case of two genes (S4 and S7), there was insufficient SNP variation between parental alleles in the exon sequences for the use of diagnostic restriction enzymes or allele-specific primers, and so we used diagnostic restriction sites within intron sequences, which meant that those CAPS markers did not work for cDNA.

Visualisation of CAPS and homoeologue-specific PCR markers

CAPS markers were digested in a reaction volume of 10 µl, with 1–2 µl of PCR template, 1 µl enzyme buffer and 2.5–10.0 units of enzyme (New England Biosciences, Ipswich, MA, USA). CAPS were visualised on 3–4% agarose gels (for example, see Figure 1). Where differences between the restriction fragments or the lengths of the allele-specific primer products were <25 bp, 4–4.5% Metaphor agarose (Lonza, Allendale, NJ, USA) gels were used to aid in their separation. Where homoeologue losses were found using diagnostic restriction sites, PCR products were sequenced and compared with sequences of the diploid parents and *T. miscellus* individuals not showing homoeologue loss. This follow-up ensured that the detected losses were not caused by point mutations at the restriction sites. In the case of homoeologue silencing events, this was not necessary as the integrity of the diagnostic restriction sites had been shown in the genomic DNA.

Results

Homoeologue loss in natural and synthetic allopolyploids of *T. miscellus*

We developed CAPS and/or homoeologue-specific PCR primers analysis in *T. miscellus* for nine homoeologue

Figure 1 Example of an analysed CAPS marker. The stained 4% metaphor agarose gel shows homoeologue loss in two *Tragopogon miscellus* individuals (lanes 13 and 16) from the Oakesdale population for gene D1 (an LRR protein kinase). The DNA fragments are restriction-digested PCR products of this gene from individual plants. The first eight lanes of the gel show *Tragopogon pratensis* individuals (Soltis and Soltis collection number 2672, individuals 1, 2, 3, 6, 7, 8, 9 and 10, respectively); lanes 9–16 show *T. miscellus* individuals (Soltis and Soltis collection number 2671, individuals 2, 3, 4, 5, 7, 8, 10 and 11, respectively); lanes 17–23 show *T. dubius* individuals (Soltis and Soltis collection number 2670, individuals 1–7, respectively); lane 24 shows a negative PCR control; lane 25 shows HyperLadder IV (Bioline, Taunton, MA, USA) markers from 100–500 bp.

pairs that seem to be orthologous to genes that are found as singleton genes in *A. thaliana* and for one homoeologue pair that seemed orthologous to a duplicate pair in *A. thaliana*. These 10 homoeologue pairs, together with putative function and the details of the methods used to analyse them, are shown in Table 2. These methods were successful using cDNA for 7 of the 10 genes (see Table 2). We could not analyse expression for one gene that did not amplify from leaf cDNA and two genes that did not have suitable SNP variation in exon regions.

We characterised a total of 1140 loci (10 homoeologous pairs in 57 plants) in genomic DNA from five natural populations of *T. miscellus* (see Table 2), finding 18 cases of homoeologue loss (1.6%; see Table 3, Figure 1). All cases of loss were confirmed by sequencing: *T. miscellus* individuals showing homoeologue loss gave sequences identical to those of one parental diploid, whereas individuals with both homoeologues showed double peaks at SNPs and illegibility of sequence downstream of indels (see Supplementary Material, Figure S1). Thus, we showed that 3.2% of homoeologue pairs in our sample of *T. miscellus* were reduced to singleton status within 80 years of polyploidisation. Of these cases, 7 involved loss of the *T. pratensis* homoeologue and 11 involved loss of the *T. dubius* homoeologue. As some of the homoeologue losses within the same population were of the same gene and homoeologue, these may have descended from a single loss event in a common ancestor. Therefore, the minimum number of independent loss events that could explain our data is 13, assuming an independent origin for each population, which is supported by recent microsatellite data suggesting that individual populations may only rarely contain plants from lineages with separate origins (Symonds *et al.*, in preparation).

The homoeologue losses were not distributed evenly across the 10 genes analysed (Table 3). Four of the genes showed no homoeologue losses, and 5 of the 18 losses occurred at a single locus (S3). The frequency of homoeologue losses varied across the five populations sampled. Eight of the losses were in the Moscow population (0.44 per individual analysed), five in the Oakesdale population (0.63 per individual), three in the Garfield population (0.38 per individual), one in the Spangle population (0.13 per individual) and one in the Pullman population (0.06 per individual). In one case (Moscow 2604-20), a single plant showed homoeologue loss in more than one gene, with the *T. dubius* copy lost in both cases (Table 4).

In 44 synthetic first-generation (S_1) *T. miscellus* plants, we found no evidence of homoeologue loss for the 10 genes investigated.

Expression of homoeologous loci in natural and synthetic allopolyploids of *T. miscellus*

We characterised the expression of 474 loci (seven homoeologous pairs in 35 plants; 16 loci were not characterised as eight PCR reactions failed to work) using cDNA from natural populations of *T. miscellus*, finding 16 cases of homoeologue silencing (3.4%; see Table 3). Thus, 6.8% of homoeologue pairs in our sample of *T. miscellus* did not express one of the homoeologues that was present in the genome. This silencing is in addition to losses of homoeologues, which occurred in 3.2% of homoeologous pairs examined (see above).

Table 3 Loss and silencing of homoeologues of 10 genes in natural populations of *Tragopogon miscellus*

Gene	<i>T. pratensis</i> homoeologues lost	<i>T. dubius</i> homoeologues lost	Homoeologues retained	Proportion lost	Minimum number of loss events	<i>T. pratensis</i> homoeologues silenced	<i>T. dubius</i> homoeologues silenced	Homoeologues expressed or lost from cDNA	Proportion silenced	Minimum number of silencing events
S2	0	0	114	0.0000	0	0	2	66	0.0294	2
S3	2	4	108	0.0526	3	3	2	65	0.0714	3
S4	0	4	110	0.0351	3	—	—	—	—	—
S5	0	0	114	0.0000	0	0	1	69	0.0142	1
S7	0	0	114	0.0000	0	—	—	—	—	—
S8	1	1	112	0.0175	2	2	1	65	0.0441	2
S13	0	0	114	0.0000	0	—	—	—	—	—
S18	1	0	113	0.0088	1	0	4	54	0.0690	4
S22	0	1	113	0.0088	1	0	1	69	0.0143	1
D1	3	1	110	0.0351	3	0	0	70	0.0000	0
Total	7	11	1122	0.0158	13	5	11	458	0.0338	13

Table 4 Individual plants showing loss or silencing of homoeologues

Population	Plant number	Loci showing loss	Loci showing silencing
Spangle	2693-14		S3 _p , S18 _d
Spangle	2693-8		S3 _p
Spangle	2693-7		S3 _p
Spangle	2693-6	S4 _d	
Pullman	2605-14		S3 _d
Pullman	2605-28	S3 _d	
Pullman	2605-13a		S18 _d
Oakesdale	2671-11	D1 _d	S2 _d
Oakesdale	2671-10	S3 _d	
Oakesdale	2671-8		S8 _p
Oakesdale	2671-7	D1 _p	
Oakesdale	2671-5	S3 _d	
Oakesdale	2671-4	S22 _d	
Oakesdale	2671-2		S8 _p
Moscow	2604-30x	S3 _p	
Moscow	2604-26a	S18 _p	
Moscow	2604-24x		S18 _d
Moscow	2604-22a	D1 _p	
Moscow	2604-21	D1 _p	
Moscow	2604-20	S3 _d , S8 _d	
Moscow	2604-17	S4 _d	
Moscow	2604-15a	S3 _p	
Garfield	2688-10	S4 _d	
Garfield	2688-6	S8 _p	
Garfield	2688-3	S4 _d	S8 _d , S5 _d , S3 _d , S2 _d , S18 _d , S22 _d

Homoeologue subscripts show parental species (d = *T. dubius*; P = *T. pratensis*).

Eleven of the homoeologues silenced were from *T. dubius*, and five were from *T. pratensis*. If silencing of genes occurred in a heritable manner (for example, Kloc *et al.*, 2008), the minimum number of heritable silencing events that must have occurred is 13 (Table 3).

The frequency of homoeologue silencing differed among the seven genes analysed, with the number of silencing events per gene varying between zero and five (Table 3). Homoeologue silencing differed in its frequency across the five natural populations of *T. miscellus* analysed. One of the silencing events was in the Moscow population (0.1 per individual analysed), two in the Pullman population (0.2 per individual), three in the Oakesdale population (0.5 per individual), four in the Spangle population (0.8 per individual) and six in the Garfield population (1.2 per individual). Two individuals showed homoeologue silencing in more than one gene: a plant from the Spangle population had two homoeologues silenced, one from *T. dubius* and one from *T. pratensis*, and one plant from the Garfield population had six homoeologues silenced, all from *T. dubius* (Table 4).

In 44 synthetic first-generation (S₁) *T. miscellus* plants, we found no evidence for silencing of homoeologues in the seven genes investigated.

Discussion

This analysis of homoeologue evolution in *T. miscellus* shows the following: (1) no genes studied showed instantaneous loss or silencing of one homoeologue in re-synthesised first-generation *T. miscellus*; (2) loss and silencing of some homoeologues begin in natural populations within 40 generations (that is, 80 years) after whole-genome duplication in the natural allopolyploid

T. miscellus; (3) none of the losses or silencing events we found was fixed within populations; (4) some loss or silencing events occurred independently in more than one population, but not with the consistency to suggest a concerted mechanism; (5) silencing of homoeologues at the loci surveyed is slightly more frequent than homoeologue loss overall.

Timing of homoeologue loss

In a survey of 10 loci in 59 plants from five natural populations of *T. miscellus*, we found 3.2% of the homoeologue pairs studied to return to singleton status in at least one individual. In a survey of 10 different loci in 20 individuals from two natural populations (Moscow and Pullman), Tate *et al.* (2006) found the rate of homoeologue loss to be higher: 10% of the homoeologue pairs surveyed. Together, these two studies concur in showing that loss of homoeologues is not uncommon within 40 generations of allopolyploidisation in natural *T. miscellus*.

Losses of homoeologues found here are not an immediate consequence of hybridisation and whole-genome duplication. If they were, we would have found them in the S₁-generation synthetic allopolyploids, and we would expect them to be fixed in the natural populations. Rather, they seem to have arisen gradually over time and to be varying in frequency in the natural populations, with some losses currently at a higher frequency than others. We found no homoeologue losses in first-generation synthetic allopolyploids, and Tate *et al.* (2006) did not find losses in synthetic, diploid F₁ hybrids between *T. dubius* and *T. pratensis*.

These patterns contrast with those we might expect on the basis of other empirical and theoretical studies. On the one hand, synthetic wheat allopolyploids show rapid gene loss occurring in the F₁ or first synthetic (S₁) allopolyploid generation (Ozkan *et al.*, 2001; Shaked *et al.*, 2001; Kashkush *et al.*, 2002; Levy and Feldman, 2004). On the other hand, point mutation models predict a slower rate of homoeologue loss (Lynch and Force, 2000). Perhaps polyploidy in *T. miscellus* sets the stage for non-instantaneous, yet rapid, mutational mechanisms that are ongoing through many generations. The most obvious of these is homoeologous recombination, in which fragments of chromosomes can be lost, though we cannot exclude the possibility of gene conversion, as has been found for rRNA genes in the species (Kovarík *et al.*, 2005; Matyasek *et al.*, 2007). Homoeologous recombination seems to have caused loss of chromosome fragments in re-synthesised *Brassica* allopolyploids (Song *et al.*, 1995; Gaeta *et al.*, 2007). Ownbey (1950) observed multivalent formation in early generations of natural *T. miscellus*, and the patterns of isozyme variation in *T. miscellus* are consistent with homoeologous recombination (Soltis *et al.*, 1995). More recently, Lim *et al.* (2008) and Tate *et al.* (in press) reported multivalent formation in both natural and synthetic *Tragopogon* allopolyploids, along with unisomy, trisomy and reciprocal translocations in natural *Tragopogon* allopolyploids. Thus, it seems that in *T. miscellus* homoeologue loss is an ongoing consequence of allopolyploidy, rather than an instantaneous event following allopolyploidisation.

We did not find a marked preferential loss of *T. dubius* versus *T. pratensis* homoeologues in this study, but if we

combine our results with those of Tate *et al.* (2006), the total number of *T. dubius* homoeologue losses is 27, and the total number of *T. pratensis* homoeologue losses is 11. This bias towards the loss of *T. dubius* homoeologues is close to statistical significance (paired *t*-test: d.f. = 19, $t = 2.027$, $P = 0.057$). A similar bias towards loss of *T. dubius* genetic material has been found for rDNA copy number (Kovarík *et al.*, 2005). This could be because of cytoplasmic–nuclear interactions. All the *T. miscellus* populations that we studied, except Pullman (in which we found only one loss), have *T. pratensis* as the maternal parent and there may be selection to keep the cytoplasmic and nuclear genomes as similar in their ancestry as possible. A bias in the loss of homoeologues has been found in synthetic polyploids of *Brassica* (Song *et al.*, 1995). In *T. mirus*, which has *T. dubius* as the paternal parent and *T. porrifolius* as the maternal parent, a bias towards loss of *T. dubius* homoeologues was also found using CAPS markers (Koh *et al.*, in preparation).

Timing of homoeologue silencing

We found no evidence for silencing in the 10 homoeologues studied in the first-generation synthetic (S_1) allopolyploids. In natural *T. miscellus*, we found silencing (loss of expression) of homoeologues to be more frequent than loss of homoeologues at the seven loci examined, and more stochastic. Six of the 16 cases of homoeologue silencing were found in a single individual (2106; see Table 4). This contrasts with the 10 loci in Tate *et al.* (2006), in which loss was more frequent than silencing. If the expression of both homoeologues at certain loci reduced plant fitness, then we might expect to see a correlation between loci where loss occurs, and loci where silencing occurs, but this was not found. It is possible that some homoeologue silencing may be specific to the rosette leaf tissue analysed here, and therefore tissue-specific expression could be occurring (Adams *et al.*, 2003, 2004; Adams and Wendel, 2005). Further work is needed to compare the expression of homoeologues in other tissues.

In natural cotton polyploids, a micro-array study of 1383 genes found biased expression in 70% of homoeologue pairs, 24% of which seem to have biased expression immediately on diploid hybridisation of the parental species (Flagel *et al.*, 2008). In synthetic *Senecio cambrensis* allopolyploids, anonymous microarrays with ~6000 cDNA clones showed that gene expression changes are extensive in diploid hybrids of the parental species, but are ameliorated by polyploidisation (Hegarty *et al.*, 2005, 2006). In synthetic allopolyploids between *A. thaliana* and *A. suecica*, a microarray study with 26090 annotated genes found that ~5% showed expression changes (Madlung *et al.*, 2005; Wang *et al.*, 2006).

The current study seems to contrast with these studies in finding no changes in gene expression in the S_1 generation. This may be for the following reasons: (1) The methods used here, which examine gene expression by the presence or absence of a cDNA band in an agarose gel, are conservative in that they are not sensitive to slight down- or upregulation of genes. Although we can detect homoeologue silencing, small differences in the level of expression between homoeologues are likely to have gone undetected. (2) *T. miscellus* is an excellent evolutionary model, but

not yet a genetic 'model organism' (Soltis *et al.*, 2004b; Buggs, 2008). Our sample size of genes is much smaller than those conducted using microarrays; so other genes not studied here may show changes in expression. (3) Allopolyploidisation may differ between species in the mode and tempo of its effects on the genome and transcriptome. One cause of such differences could be the degree of relatedness between the progenitor species that hybridise to give rise to an allopolyploid: loss of genes through homoeologous recombination may be more likely to occur when parental species are more closely related, whereas distantly related parents may have chromosomes that are differentiated enough not to recombine (Darlington, 1937; Clausen *et al.*, 1945; Buggs *et al.*, 2008). On the other hand, hybridisation between distantly related species might only produce viable offspring if genomic and transcriptomic re-adjustments take place immediately (for example, Ozkan *et al.*, 2001). The degree to which genes are silenced in parental species may be another cause of varied consequences of allopolyploidisation among taxa, as epigenetic silencing may be disrupted by allopolyploidisation (McClintock, 1984; Comai, 2000; Comai *et al.*, 2003).

Conclusion

This study and Tate *et al.* (2006) provide the first evaluation of homoeologue loss and silencing in recent natural polyploids, in comparison with F_1 hybrids and first-generation synthetic allopolyploids of the same species. We show the dynamic nature of the early generations after allopolyploidisation, which, in *T. miscellus*, seem to have greater consequences for homoeologue silencing and loss than hybridisation and genome doubling do. New pyrosequencing technologies will enable the development of far more extensive genomic resources in *Tragopogon*, allowing us to follow up these intriguing results with a survey of many hundreds of homoeologue sets.

Acknowledgements

We thank Andrew Leitch, Jonathan Wendel and two anonymous reviewers for helpful comments on earlier versions of this paper. We also thank Bryon Peterson and Ralph Brooks for assistance in the field. Funding for this research was provided by the University of Florida and NSF Grants DEB-0614421 to Doug and Pam Soltis, MCB-0346437 to Doug and Pam Soltis and Jennifer Tate, and MCB-0450260 to Andrew Paterson.

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. *Proc Natl Acad Sci 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.
- Adams KL, Wendel JF (2005). Allele-specific, bidirectional silencing of an alcohol dehydrogenase gene in different organs of interspecific diploid cotton hybrids. *Genetics* **171**: 2139–2142.
- Buggs RJA (2008). Towards natural polyploid model organisms. *Mol Ecol* **17**: 1875–1876.

- Buggs RJA, Soltis PS, Mavrodiev EV, Symonds VV, Soltis DE (2008). Does phylogenetic distance between parental genomes govern the success of polyploids? *Castanea* **73**: 74–93.
- Chen LZ, Lou QF, Zhuang Y, Chen JF, Zhang XQ, Wolukau JN (2007). Cytological diploidization and rapid genome changes of the newly synthesized allotetraploids *Cucumis* × *hytivus*. *Planta* **225**: 603–614.
- Clausen J, Keck DD, Hiesey WM (1945). Experimental studies on the nature of species. II. Plant evolution through amphiploidy and autopoloidy, with examples from the Madiinae. *Carnegie Inst Wash Publ* 564.
- Comai L (2000). Genetic and epigenetic interactions in allopolyploid plants. *Plant Mol Biol* **43**: 387–399.
- Comai L (2005). The advantages and disadvantages of being polyploid. *Nat Rev Genet* **6**: 836–846.
- Comai L, Madlung A, Josefsson C, Tyagi A (2003). Do the different parental 'heteromes' cause genomic shock in newly formed allopolyploids? *Philos Trans R Soc Lond B Biol Sci* **358**: 1149–1155.
- Darlington CD (1937). *Recent Advances in Cytology*. Blakiston: Philadelphia.
- Douglas SE (1995). DNA strider. *Mol Biotechnol* **3**: 37–45.
- Doyle J, Doyle JL (1987). Genomic plant DNA preparation from fresh tissue-CTAB method. *Phytochem Bull* **19**: 11–15.
- Flagel LE, Udall J, Nettleton D, Wendel J (2008). Duplicate gene expression in allopolyploid *Gossypium* reveals two temporally distinct phases of expression evolution. *BMC Biol* **6**: 16.
- Gadberry MD, Malcomber ST, Doust AN, Kellogg EA (2005). Primaclade—a flexible tool to find conserved PCR primers across multiple species. *Bioinformatics* **21**: 1263–1264.
- 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 J* **50**: 995–1006.
- Hegarty M, Barker G, Wilson I, Abbott RJ, Edwards KJ, Hiscock SJ (2006). Transcriptome shock after interspecific hybridization in *Senecio* is ameliorated by genome duplication. *Curr Biol* **16**: 1652–1659.
- Hegarty M, Jones J, Wilson I, Barker G, JA C, Sanchez-Baracaldo P *et al.* (2005). Development of anonymous cDNA microarrays to study changes to the *Senecio* floral transcriptome during hybrid speciation. *Mol Ecol* **14**: 2493–2510.
- Hegarty MJ, Hiscock SJ (2008). Genomic clues to the evolutionary success of polyploid plants. *Curr Biol* **18**: R435–R444.
- Kashkush K, Feldman M, Levy AA (2002). Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. *Genetics* **160**: 1651–1659.
- Kloc A, Zaratiegui M, Nora E, Martienssen R (2008). RNA interference guides histone modification during the S phase of chromosomal replication. *Curr Biol* **18**: 490–495.
- Konieczny A, Ausubel FM (1993). A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. *Plant J* **4**: 403–410.
- Kovarik A, Pires JC, Leitch AR, Lim KY, Sherwood AM, Matyasek R *et al.* (2005). Rapid concerted evolution of nuclear ribosomal DNA in two *Tragopogon* allopolyploids of recent and recurrent origin. *Genetics* **169**: 931–944.
- Leitch AR, Leitch IJ (2008). Genomic plasticity and the diversity of polyploid plants. *Science* **320**: 481–483.
- Leitch IJ, Bennett MD (2004). Genome downsizing in polyploid plants. *Biol J Linn Soc* **82**: 651–663.
- Levy AA, Feldman M (2004). Genetic and epigenetic reprogramming of the wheat genome upon allopolyploidization. *Biol J Linn Soc* **82**: 607–613.
- Lim KY, Soltis DE, Soltis PS, Tate JA, Matyasek R, Srubarova H *et al.* (2008). Rapid chromosome evolution in recently formed polyploids in *Tragopogon* (Asteraceae). *PLoS ONE* **3**: e3353.
- Liu B, Vega JM, Feldman M (1998). Rapid genomic changes in newly synthesized amphiploids of *Triticum* and *Aegilops*. II. Changes in low-copy coding DNA sequences. *Genome* **41**: 535–542.
- Lynch M, Force A (2000). The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**: 459–473.
- Maddison J, Maddison M (2004). *McClade Version 4.06*. Sinauer Associates, Inc.: Sunderland, Mass.
- Madlung A, Tyagi AP, Watson B, Jiang HM, Kagochi T, Doerge RW *et al.* (2005). Genomic changes in synthetic *Arabidopsis* polyploids. *Plant J* **41**: 221–230.
- Matyasek R, Tate JA, Lim YK, Srubarova H, Koh J, Leitch AR *et al.* (2007). Concerted evolution of rDNA in recently formed *tragopogon* allotetraploids is typically associated with an inverse correlation between gene copy number and expression. *Genetics* **176**: 2509–2519.
- Mavrodiev EV, Tancig M, Sherwood AM, Gitzendanner MA, Rocca J, Soltis PS *et al.* (2005). Phylogeny of *Tragopogon* L (Asteraceae) based on internal and external transcribed spacer sequence data. *Int J Plant Sci* **166**: 117–133.
- McClintock B (1984). The significance of responses of the genome to challenge. *Science* **226**: 792–801.
- Novak SJ, Soltis DE, Soltis PS (1991). Ownbey *Tragopogons*—40 years later. *Am J Bot* **78**: 1586–1600.
- Ownbey M (1950). Natural hybridization and amphiploidy in the genus *Tragopogon*. *Am J Bot* **37**: 487–499.
- Ozkan H, Feldman M (2001). Genotypic variation in tetraploid wheat affecting homoeologous pairing in hybrids with *Aegilops peregrina*. *Genome* **44**: 1000–1006.
- Ozkan H, Levy AA, Feldman M (2001). Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. *Plant Cell* **13**: 1735–1747.
- Ozkan H, Levy AA, Feldman M (2002). Rapid differentiation of homeologous chromosomes in newly-formed allopolyploid wheat. *Isr J Plant Sci* **50**: S65–S76.
- Paterson AH, Chapman BA, Kissinger JC, Bowers JE, Feltus FA, Estill JC (2006). Many gene and domain families have convergent fates following independent whole-genome duplication events in *Arabidopsis*, *Oryza*, *Saccharomyces* and *Tetraodon*. *Trends Genet* **22**: 597–602.
- Rozen S, Skaletsky H (1999). Primer3 on the WWW for general users and for biologist programmers. In: Misener S and Krawetz SA (eds). *Methods in Molecular Biology*. Humana Press: Totowa, NJ. Vol. 132: *Bioinformatics Methods and Protocols*, pp 365–386.
- 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.
- Soltis DE, Soltis PS, Pires JC, Kovarik A, Tate JA, Mavrodiev E (2004a). Recent and recurrent polyploidy in *Tragopogon* (Asteraceae): cytogenetic, genomic and genetic comparisons. *Biol J Linn Soc* **82**: 485–501.
- Soltis DE, Soltis PS, Tate JA (2004b). Advances in the study of polyploidy since plant speciation. *New Phytol* **161**: 173–191.
- Soltis PS, Plunkett GM, Novak SJ, Soltis DE (1995). Genetic variation in *Tragopogon* species: additional origins of the allotetraploids *T. mirus* and *T. miscellus* (Compositae). *Am J Bot* **82**: 1329–1341.
- Song K, Lu P, Tang K, Osborn TC (1995). Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. *Proc Natl Acad Sci USA* **92**: 7719–7723.
- Tate JA, Ni ZF, Scheen AC, Koh J, Gilbert CA, Lefkowitz D *et al.* (2006). Evolution and expression of homeologous loci in

- Tragopogon miscellus* (Asteraceae), a recent and reciprocally formed allopolyploid. *Genetics* **173**: 1599–1611.
- Tate JA, Symonds VV, Doust AN, Buggs RJA, Mavrodiev E, Majure LC *et al.* (in press). Synthetic polyploids of *Tragopogon miscellus* *T. mirus* (Asteraceae): 60 years after Ownbey's discovery. *Am J Bot.*
- Wang J, Tian L, Lee H-S, Wei NE, Jiang H, Watson B *et al.* (2006). Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. *Genetics* **172**: 507–517.
- Wang JL, Tian L, Madlung A, Lee HS, Chen M, Lee JJ *et al.* (2004). Stochastic and epigenetic changes of gene expression in *Arabidopsis* polyploids. *Genetics* **167**: 1961–1973.
- Wolfe KH (2001). Yesterday's polyploids and the mystery of diploidization. *Nat Rev Genet* **2**: 333–341.
- Ye S, Dhillon S, Ke X, Collins AR, Day INM (2001). An efficient procedure for genotyping single nucleotide polymorphisms. *Nucl Acids Res* **29**: e88.

Supplementary information accompanies the paper on Heredity website (<http://www.nature.com/hdy>)