PLOS | ONE

RESEARCH ARTICLE

Cytonuclear Coordination Is Not Immediate upon Allopolyploid Formation in *Tragopogon miscellus* (Asteraceae) Allopolyploids

Tina Sehrish[1], V. Vaughan Symonds[1], Douglas E. Soltis[2,3,4], Pamela S. Soltis[3,4], Jennifer A. Tate[1]*

1 Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand, **2** Department of Biology, University of Florida, Gainesville, Florida, United States of America, **3** Florida Museum of Natural History, University of Florida, Gainesville, Florida, United States of America, **4** Genetics Institute, University of Florida, Gainesville, Florida, United States of America

* j.tate@massey.ac.nz

Citation: Sehrish T, Symonds VV, Soltis DE, Soltis PS, Tate JA (2015) Cytonuclear Coordination Is Not Immediate upon Allopolyploid Formation in *Tragopogon miscellus* (Asteraceae) Allopolyploids. PLoS ONE 10(12): e0144339. doi:10.1371/journal.pone.0144339

Editor: Khalil Kashkush, Ben-Gurion University, ISRAEL

Received: October 15, 2015

Accepted: November 17, 2015

Published: December 8, 2015

Data Availability Statement: Sequence data have been deposited to GenBank under accession numbers: KT897488-KT897491 and KT879189-KT879190.

Funding: A PhD scholarship to TS from the Higher Education Commission of Pakistan funded TS. Massey University Research Fund to JT funded the research. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Allopolyploids, formed by hybridization and chromosome doubling, face the immediate challenge of having duplicated nuclear genomes that interact with the haploid and maternally inherited cytoplasmic (plastid and mitochondrial) genomes. Most of our knowledge of the genomic consequences of allopolyploidy has focused on the fate of the duplicated nuclear genes without regard to their potential interactions with cytoplasmic genomes. As a step toward understanding the fates of nuclear-encoded subunits that are plastid-targeted, here we examine the retention and expression of the gene encoding the small subunit of Ribulose-1, 5-bisphosphate carboxylase/oxygenase (Rubisco; *rbcS*) in multiple populations of allotetraploid *Tragopogon miscellus* (Asteraceae). These polyploids formed recently (~80 years ago) and repeatedly from *T*. *dubius* and *T*. *pratensis* in the northwestern United States. Examination of 79 *T*. *miscellus* individuals from 10 natural populations, as well as 25 synthetic allotetraploids, including reciprocally formed plants, revealed a low percentage of naturally occurring individuals that show a bias in either gene (homeolog) loss (12%) or expression (16%), usually toward maintaining the maternal nuclear copy of *rbcS*. For individuals showing loss, seven retained the maternally derived *rbcS* homeolog only, while three had the paternally derived copy. All of the synthetic polyploid individuals examined (S_0 and S_1 generations) retained and expressed both parental homeologs. These results demonstrate that cytonuclear coordination does not happen immediately upon polyploid formation in *Tragopogon miscellus*.

Introduction

Allopolyploidy is a major mode of plant speciation and results from the union of two or more diverse, but generally closely related, genomes by hybridization and genome duplication [1, 2]. Genomic data indicate that all angiosperms may be regarded as polyploid, if paleopolyploid events are taken into account [3–5]. Allopolyploid genomes experience both immediate and

Competing Interests: The authors have declared that no competing interests exist.

long-term evolutionary changes, which may involve a variety of genetic and epigenetic interactions leading to genome alteration, regulatory incompatibilities, chromosomal abnormalities, and reproductive challenges [6–14]. Polyploidy has been considered a driver of modifications in gene function, potentially resulting in four fates for the duplicated genes (homeologs): (I) both copies are preserved and retain their original functions, (II) one copy maintains the original function whereas the other copy is silenced, or (III) the two copies diverge such that each copy contributes only a part of the original gene function (subfunctionalization) or (IV) one copy attains a novel function (neofunctionalization) [15–21].

In newly formed allopolyploids, coordination between the haploid maternally inherited cytoplasmic (plastid and mitochondrial) and the duplicated biparentally inherited nuclear genomes is required to facilitate genomic stability [8]. Indeed, 'cytonuclear interactions' are considered responsible for post-zygotic hybrid incompatibilities and speciation [22–24] and have also caused striking differences in floral traits in reciprocal diploid hybrids [25, 26]. Cytonuclear coordination may also be a contributor to the directional genomic changes and preferential expression of some genes in reciprocally formed polyploids [27–29].

Recent studies of Rubisco (Ribulose-1,5-bisphosphate carboxylase/oxygenase), which comprises a nuclear-encoded subunit (rbcS) and a chloroplast-encoded subunit (rbcL), in allopolyploids have revealed a dynamic nature to the evolution of the nuclear component [30, 31]. In several allopolyploid systems, the duplicated nuclear gene copies (homeologs) of *rbcS* undergo gene conversion in favor of maintaining the maternally derived copy even when the parental chloroplast sequences of *rbcL* are not divergent [30]. Additionally, a common feature of this system is that allopolyploids show preferential expression of the maternal *rbcS* homeolog when both copies are maintained in the genome [30, 31]. How early following polyploid formation this cytonuclear coordination might be established is not known as the polyploids studied to date are several hundred thousand to several million years old [30, 31].

An excellent model system for studying the early stages of allopolyploid cytonuclear coordination is offered by *Tragopogon* (Asteraceae). Following the introduction of three diploid species from Europe (*Tragopogon dubius*, *T. pratensis*, and *T. porrifolius*) to the Palouse region of eastern Washington State/western Idaho, USA, in the early 1900s, two allopolyploid species were formed. *Tragopogon mirus* (*T. dubius* × *T. porrifolius*) and *T. miscellus* (*T. dubius* × *T. pratensis*) both formed repeatedly in the past 80 years in western North America with *T. miscellus* also forming reciprocally, yielding short-liguled (*T. dubius* ♂ × *T. pratensis* ♀) and long-liguled (*T. dubius* ♀ × *T. pratensis* ♂) forms [32–34]. Recurrent formation of both allopolyploids and restricted gene flow among origins [35–38] offer an opportunity to determine if independently formed polyploids develop similar cytonuclear coordination. Previous studies have identified a myriad of genomic and transcriptomic modifications in the *Tragopogon* allopolyploids in the short time since their formation, including differential expression of homeologous loci, homeolog loss and silencing, differential proteomes [1, 39–46], and extensive chromosomal variation, such as aneuploidy and intergenomic translocations [13, 47]. Moreover, the formation of synthetic polyploids of *Tragopogon* has allowed the analysis of genomic modifications at early stages of polyploid formation [48].

Here, we use the Rubisco system (*rbcS* and *rbcL*) to examine cytonuclear coordination in naturally occurring and synthetic *Tragopogon miscellus* allopolyploids, representing independent and reciprocal formations. We characterize *rbcS* in the *Tragopogon* diploid parental species to answer the following questions: (1) How divergent are *rbcS* and *rbcL* in the *T. miscellus* progenitors? (2) Is there differential retention of *rbcS* homeologs in *T. miscellus*? (3) When both parental copies of *rbcS* are retained, do the naturally occurring and synthetic polyploids of *T. miscellus* show equal or biased expression of the *rbcS* homeologs?

Materials and Methods

Plant material

The populations sampled for *Tragopogon dubius* Scop. *T. pratensis* L. and *T. miscellus* Ownbey are listed in S1 Table, as are the synthetic lineages of *T. miscellus* examined. For *T. miscellus*, we included individuals from nine short-liguled populations (*T. pratensis* maternal parent) and one long-liguled population (*T. dubius* maternal parent), the latter representing the only extant natural population of this form. To assess potential variability in the diploid progenitors, genomic DNA and cDNA were included for multiple individuals of *T. dubius* (12) and *T. pratensis* (8) from different populations (S1 Table). For *T. miscellus*, four synthetic lineages (25 individuals) and ten populations (79 individuals total) were sampled. Plant material for most of the polyploids and diploids was the same as that used in Tate *et al.* [45, 46] and Buggs *et al.* [40]. For the synthetics, mature seeds were grown under standard glasshouse conditions at Massey University (Palmerston North, New Zealand); these lines were generated by Tate *et al.* [48]. For expression analyses, only a subset of individuals (31) was studied due to limited availability of fresh material for RNA extraction (S1 Table).

DNA and RNA extraction

Both DNA and RNA were extracted from leaf tissue 28 days after seed germination. For DNA, a modified CTAB extraction protocol was used [49]. For RNA extraction, leaf tissue was flash-frozen in liquid nitrogen and ground in a 1.5-ml tube using a sterile pestle. Total RNA was extracted using the RNeasy Plant Mini kit (Qiagen, UK). First-strand cDNA was synthesized from 200 ng of total RNA using SuperScript III First-Strand Synthesis System for RT-PCR (Invitrogen, CA, USA).

Primer design, PCR and sequencing of *rbcL* and *rbcS-1*

Full length *rbcL* was amplified from the diploid progenitors and *T. miscellus* using primers *rbcL1* and *rbcL2* ([50, 51] primer names as in [52]). PCR reactions were conducted in a 25-μl total volume containing 10X Thermopol buffer (New England Biolabs, USA), 10 mM dNTPs, 5 μM each primer, 0.5 Unit NEB *Taq* polymerase and ~50 ng of either genomic DNA or cDNA template. The following PCR profile was used: 95°C for 5 min, 48°C for 45 sec, 72°C for 1 min followed by 35 cycles at 95°C for 1 min, 48°C for 45 sec (2 sec added in each successive cycle) and 72°C for 1 min, with a final extension at 72°C for 10 min [52].

Initial amplification of *rbcS* was accomplished by designing PCR primers from an alignment of *T. dubius* ESTs (Tdu01-5MS1_K18.e, Tdu01-3MS1_B10.e, Tdu01-2Ms1_K16.e) to *Lactuca sativa rbcS* (AF162210) using Primer3 [53]. Primers (*rbcS*-2F and *rbcS*-2R) for one copy, hereafter *rbcS-1*, were used for the initial amplification of both genomic and cDNA of the diploids (Table 1). A second *rbcS*-like sequence was identified in the *T. dubius* EST database, but this copy is apparently a pseudo-gene as it is truncated (missing 5' UTR through exon 1) with several premature stop codons and indels as compared to the full-length *rbcS-1* and *rbcS* sequences from other Asteraceae (S1 Fig). For this second copy, *rbcS-2*, 5' genome walking (using methods described later for *rbcS-1*) revealed the presence of a long ~700-bp intron-like sequence (data not shown) that has not been found in any angiosperm group to date [54]. Likewise, this *rbcS-2* copy is also truncated in *T. pratensis*. Amplification of *rbcS-1* was conducted using the following PCR profile: 95°C for 5 min, 95°C for 1 min, 53°C for 1 min, 72°C for 1 min for 5 cycles, followed by 44 cycles of 95°C for 1 min, 48°C for 1 min, 72°C for 1 min and a final extension at 72°C for 7 min. PCR products of *rbcS-1* from genomic DNA of *T. dubius* and *T. pratensis* were cloned using the TOPO TA Cloning Kit (Invitrogen, CA, USA).

Table 1. *rbcS-1* primers designed in this study.

Primer Name	Experiment(s)	Primer/oligo Sequence (5'to 3')
GS1	5' Genome walking	ATCATACCTTCATGCACTGCACTCTTCCAC
GS2	5' Genome walking	AGGAAAAGTCATTGGCCTTCTTGGTGACTG
AP1	5' Genome walking	GTAATTCGCATCACTATAGCTC
AP2	5' Genome walking	ACTATAGCTCACCGCTGGT
NA44	5' Genome walking	GTAATTCGCATCACTATAGCTCACCGCTGGTCGACGGCCCGGGCTGGT
NA45	5' Genome walking	PO4-ACCAGCCC-NH2
Inv. Fwd 1	3' RACE	TGGACCTCAATCGGGTTTAT
Inv. Fwd 2	3' RACE	CAAGAAGGAGTACCCCAACG
3'RACE adapter	3' RACE	GACTCGAGTCGACATCG
3'RACE oligodT adapter	3' RACE	GACTCGAGTCGACATCGATTTTTTTTTTTTTTTTV
rbcS-2F	Sequencing and CAPS	AATGGCTTCCATCTCCTCCT
rbcS-3F	Sequencing and CAPS	TTTCCCAGTCACCAAGAAGG
rbcS-2R	Sequencing and CAPS	AGGCAACTTCCACATTGTCC
rbcS-8R	Sequencing and CAPS	CGATTGAGGTCCATCCAAAG
rbcS-F1	Sequencing	CAAAACATACCCATAACGTATCAGCC
rbcS-R3	Sequencing	AGCAGAAACATAAATTTTTATTATTATCATC
HS-Pra-Snp3	Homeolog-specific RT-PCR (forward)	AAGGCCAATGACTTTTCCTCCCGC
HS-Dub-Snp3A	Homeolog-specific RT-PCR (forward)	AAGGCCAATGACTTTTCCTCCCAT
HS-R3	Homeolog-specific RT-PCR (reverse)	CGAACATAGGCAACTTCCACATTGTCC

doi:10.1371/journal.pone.0144339.t001

Ten positive clones per sample were sequenced. Prior to sequencing, PCR products were treated with exonuclease I (5 Units) and shrimp alkaline phosphatase (0.5 Unit). Cycle sequencing was performed using Big Dye v.3.1 (Applied Biosystems, Inc.), and purified products were sequenced on an ABI DNA Analyzer 3770 at the Massey Genome Service (Palmerston North, New Zealand) using both T3 and T7 plasmid primers. Sequencing results were analyzed in Sequencher v.5.1 (Gene Codes Corporation, Michigan, USA). Based on the alignment of these cloned sequences with available *T. dubius* ESTs, a new reverse primer (*rbcS*-8R) was designed further downstream to amplify a longer portion of *rbcS-1* from synthetic and naturally occurring *T. miscellus* polyploids; these longer fragments of *rbcS-1* were then sequenced directly using the aforementioned sequencing protocol with both forward (*rbcS*-2F) and reverse (*rbcS*-8R) primers (Table 1).

Genomic and cDNA CAPS analysis

Sequences of *rbcS-1* for the diploid parents were aligned to determine sequence variation that could differentiate parental homeologs in *T. miscellus*. The programs dCAPS Finder 2.0 [55] and NEB Cutter v.1.0 [56] were used to identify diagnostic restriction sites between parental *rbcS-1* sequences. Genomic and cDNA cleaved amplified polymorphic sequence (CAPS) analyses were performed for *rbcS-1* using the forward primer *rbcS*-3F and reverse primer *rbcS*-8R (Table 1). The amplified region included exon 1 (from aligned position 433 bp), intron 1 and exon 2 (to position 1054 bp) (Fig 1). The resulting PCR products from *T. dubius* and *T. pratensis* were 462 bp for cDNA and 622 bp (*T. dubius*) and 628 bp (*T. pratensis*) from genomic DNA. PCR products were digested with *MseI*, which cuts the cDNA of *T. dubius* at one position (resulting in fragment sizes 375 bp and 87 bp) and does not cut *T. pratensis*. For genomic DNA, *T. dubius* is cut at three positions (resulting in fragment sizes 272 bp, 167 bp, 154 bp and 29 bp), while *T. pratensis* is cut at two positions (resulting in

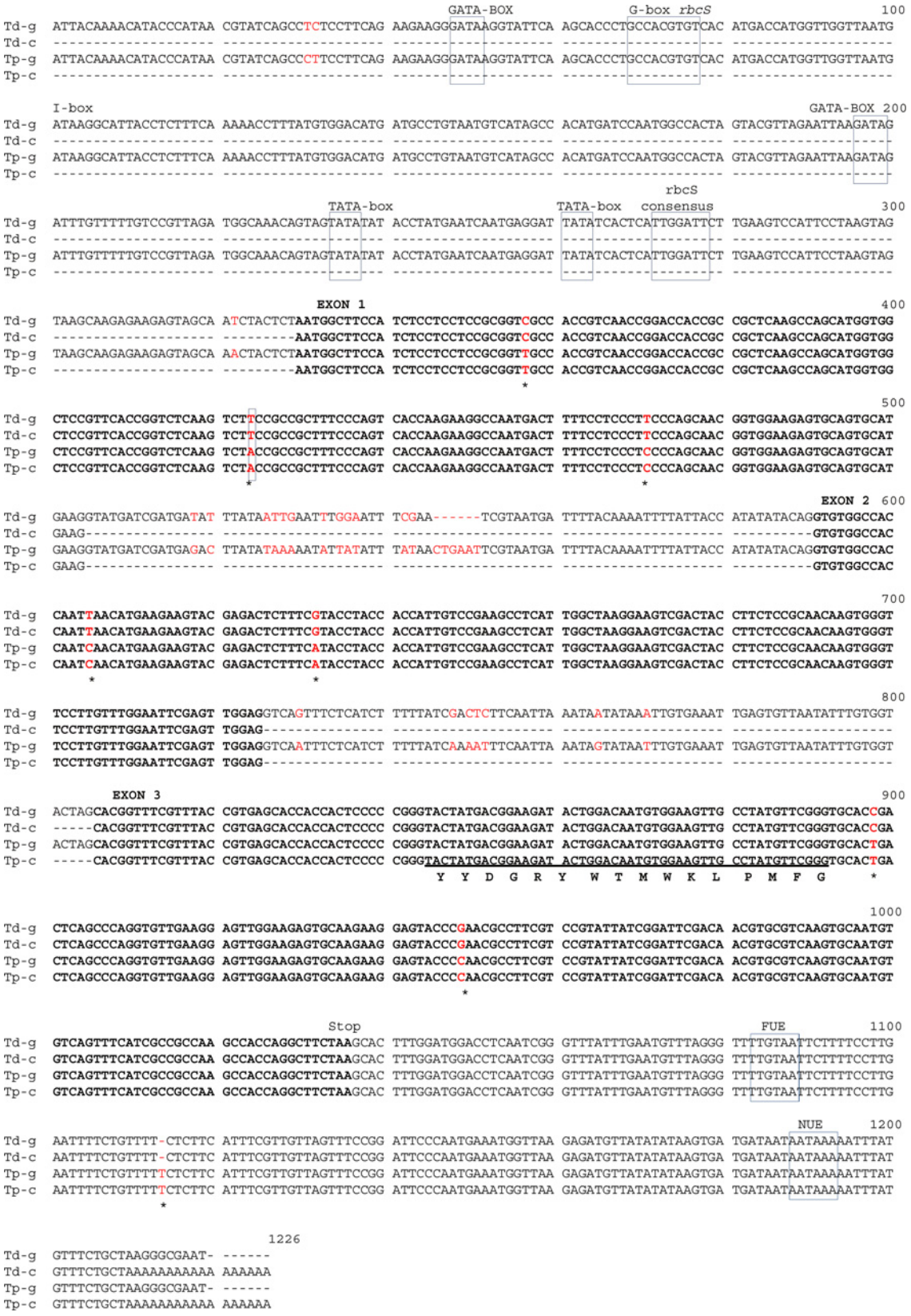
GATA-BOX
G-box rbcS
100
Td-g ATTACAAAACATACCCATAA CGTATCAGCCTCTCCTTCAG AAGAAGGGATAAGGTATTCA AGCACCCTGCCACGTGTCAC ATGACCATGGTTGGTTAATG
Td-c
Tp-g ATTACAAAACATACCCATAA CGTATCAGCCCTTCCTTCAG AAGAAGGGATAAGGTATTCA AGCACCCTGCCACGTGTCAC ATGACCATGGTTGGTTAATG
Tp-c
I-box
GATA-BOX 200
Td-g ATAAGGCATTACCTCTTTCA AAAACCTTTATGTGGACATG ATGCCTGTAATGTCATAGCC ACATGATCCAATGGCCACTA GTACGTTAGAATTAAGATAG
Td-c
Tp-g ATAAGGCATTACCTCTTTCA AAAACCTTTATGTGGACATG ATGCCTGTAATGTCATAGCC ACATGATCCAATGGCCACTA GTACGTTAGAATTAAGATAG
Tp-c
rbcS
TATA-box
TATA-box
consensus
300
Td-g ATTTGTTTTTGTCCGTTAGA TGGCAAACAGTAGTATATAT ACCTATGAATCAATGAGGAT TATATCACTCATTGGATTCT TGAAGTCCATTCCTAAGTAG
Td-c
Tp-g ATTTGTTTTTGTCCGTTAGA TGGCAAACAGTAGTATATAT ACCTATGAATCAATGAGGAT TATATCACTCATTGGATTCT TGAAGTCCATTCCTAAGTAG
Tp-c
EXON 1
400
Td-g TAAGCAAGAGAAGAGTAGCA ATCTACTCTAATGGCTTCCA TCTCCTCCTCCGCGGTCGCC ACCGTCAACCGGACCACCGC CGCTCAAGCCAGCATGGTGG
Td-c ---------AATGGCTTCCA TCTCCTCCTCCGCGGTCGCC ACCGTCAACCGGACCACCGC CGCTCAAGCCAGCATGGTGG
Tp-g TAAGCAAGAGAAGAGTAGCA AACTACTCTAATGGCTTCCA TCTCCTCCTCCGCGGTTGCC ACCGTCAACCGGACCACCGC CGCTCAAGCCAGCATGGTGG
Tp-c ---------AATGGCTTCCA TCTCCTCCTCCGCGGTTGCC ACCGTCAACCGGACCACCGC CGCTCAAGCCAGCATGGTGG
500
Td-g CTCCGTTCACCGGTCTCAAG TCTTCCGCCGCTTTCCCAGT CACCAAGAAGGCCAATGACT TTTCCTCCCTTCCCAGCAAC GGTGGAAGAGTGCAGTGCAT
Td-c CTCCGTTCACCGGTCTCAAG TCTTCCGCCGCTTTCCCAGT CACCAAGAAGGCCAATGACT TTTCCTCCCTTCCCAGCAAC GGTGGAAGAGTGCAGTGCAT
Tp-g CTCCGTTCACCGGTCTCAAG TCTACCGCCGCTTTCCCAGT CACCAAGAAGGCCAATGACT TTTCCTCCCTCCCCAGCAAC GGTGGAAGAGTGCAGTGCAT
Tp-c CTCCGTTCACCGGTCTCAAG TCTACCGCCGCTTTCCCAGT CACCAAGAAGGCCAATGACT TTTCCTCCCTCCCCAGCAAC GGTGGAAGAGTGCAGTGCAT
EXON 2
600
Td-g GAAGGTATGATCGATGATAT TTATAATTGAATTTGGAATT TCGAA------TCGTAATGA TTTTACAAAATTTTATTACC ATATATACAGGTGTGGCCAC
Td-c GAAG---------- ---------- ---------- ---------- ----------GTGTGGCCAC
Tp-g GAAGGTATGATCGATGAGAC TTATATAAAAATATTATATT TATAACTGAATTCGTAATGA TTTTACAAAATTTTATTACC ATATATACAGGTGTGGCCAC
Tp-c GAAG---------- ---------- ---------- ---------- ----------GTGTGGCCAC
700
Td-g CAATTAACATGAAGAAGTAC GAGACTCTTTCGTACCTACC ACCATTGTCCGAAGCCTCAT TGGCTAAGGAAGTCGACTAC CTTCTCCGCAACAAGTGGGT
Td-c CAATTAACATGAAGAAGTAC GAGACTCTTTCGTACCTACC ACCATTGTCCGAAGCCTCAT TGGCTAAGGAAGTCGACTAC CTTCTCCGCAACAAGTGGGT
Tp-g CAATCAACATGAAGAAGTAC GAGACTCTTTCATACCTACC ACCATTGTCCGAAGCCTCAT TGGCTAAGGAAGTCGACTAC CTTCTCCGCAACAAGTGGGT
Tp-c CAATCAACATGAAGAAGTAC GAGACTCTTTCATACCTACC ACCATTGTCCGAAGCCTCAT TGGCTAAGGAAGTCGACTAC CTTCTCCGCAACAAGTGGGT
800
Td-g TCCTTGTTTGGAATTCGAGT TGGAGGTCAGTTTCTCATCT TTTTATCGACTCTTCAATTA AATAATATAAATTGTGAAAT TGAGTGTTAATATTTGTGGT
Td-c TCCTTGTTTGGAATTCGAGT TGGAG---------- ---------- ---------- ----------
Tp-g TCCTTGTTTGGAATTCGAGT TGGAGGTCAATTTCTCATCT TTTTATCAAAATTTCAATTA AATAGTATAATTTGTGAAAT TGAGTGTTAATATTTGTGGT
Tp-c TCCTTGTTTGGAATTCGAGT TGGAG---------- ---------- ---------- ----------
EXON 3
900
Td-g ACTAGCACGGTTTCGTTTAC CGTGAGCACCACCACTCCCC CGGGTACTATGACGGAAGAT ACTGGACAATGTGGAAGTTG CCTATGTTCGGGTGCACCGA
Td-c -----CACGGTTTCGTTTAC CGTGAGCACCACCACTCCCC CGGGTACTATGACGGAAGAT ACTGGACAATGTGGAAGTTG CCTATGTTCGGGTGCACCGA
Tp-g ACTAGCACGGTTTCGTTTAC CGTGAGCACCACCACTCCCC CGGGTACTATGACGGAAGAT ACTGGACAATGTGGAAGTTG CCTATGTTCGGGTGCACTGA
Tp-c -----CACGGTTTCGTTTAC CGTGAGCACCACCACTCCCC CGGGTACTATGACGGAAGAT ACTGGACAATGTGGAAGTTG CCTATGTTCGGGTGCACTGA
Y Y D G R Y W T M W K L P M F G
1000
Td-g CTCAGCCCAGGTGTTGAAGG AGTTGGAAGAGTGCAAGAAG GAGTACCCGAACGCCTTCGT CCGTATTATCGGATTCGACA ACGTGCGTCAAGTGCAATGT
Td-c CTCAGCCCAGGTGTTGAAGG AGTTGGAAGAGTGCAAGAAG GAGTACCCGAACGCCTTCGT CCGTATTATCGGATTCGACA ACGTGCGTCAAGTGCAATGT
Tp-g CTCAGCCCAGGTGTTGAAGG AGTTGGAAGAGTGCAAGAAG GAGTACCCCAACGCCTTCGT CCGTATTATCGGATTCGACA ACGTGCGTCAAGTGCAATGT
Tp-c CTCAGCCCAGGTGTTGAAGG AGTTGGAAGAGTGCAAGAAG GAGTACCCCAACGCCTTCGT CCGTATTATCGGATTCGACA ACGTGCGTCAAGTGCAATGT
Stop
FUE
1100
Td-g GTCAGTTTCATCGCCGCCAA GCCACCAGGCTTCTAAGCAC TTTGGATGGACCTCAATCGG GTTTATTTGAATGTTTAGGG TTTTGTAATTCTTTTCCTTG
Td-c GTCAGTTTCATCGCCGCCAA GCCACCAGGCTTCTAAGCAC TTTGGATGGACCTCAATCGG GTTTATTTGAATGTTTAGGG TTTTGTAATTCTTTTCCTTG
Tp-g GTCAGTTTCATCGCCGCCAA GCCACCAGGCTTCTAAGCAC TTTGGATGGACCTCAATCGG GTTTATTTGAATGTTTAGGG TTTTGTAATTCTTTTCCTTG
Tp-c GTCAGTTTCATCGCCGCCAA GCCACCAGGCTTCTAAGCAC TTTGGATGGACCTCAATCGG GTTTATTTGAATGTTTAGGG TTTTGTAATTCTTTTCCTTG
NUE
1200
Td-g AATTTTCTGTTTT-CTCTTC ATTTCGTTGTTAGTTTCCGG ATTCCCAATGAAATGGTTAA GAGATGTTATATATAAGTGA TGATAATAATAAAAATTTAT
Td-c AATTTTCTGTTTT-CTCTTC ATTTCGTTGTTAGTTTCCGG ATTCCCAATGAAATGGTTAA GAGATGTTATATATAAGTGA TGATAATAATAAAAATTTAT
Tp-g AATTTTCTGTTTTTCTCTTC ATTTCGTTGTTAGTTTCCGG ATTCCCAATGAAATGGTTAA GAGATGTTATATATAAGTGA TGATAATAATAAAAATTTAT
Tp-c AATTTTCTGTTTTTCTCTTC ATTTCGTTGTTAGTTTCCGG ATTCCCAATGAAATGGTTAA GAGATGTTATATATAAGTGA TGATAATAATAAAAATTTAT
1226
Td-g GTTTCTGCTAAGGGCGAAT- ------
Td-c GTTTCTGCTAAAAAAAAAAA AAAAAA
Tp-g GTTTCTGCTAAGGGCGAAT- ------
Tp-c GTTTCTGCTAAAAAAAAAAA AAAAAA

Fig 1. *rbcS-1* gene structure and locations of SNPs between *Tragopogon dubius* and *T. pratensis*. Upstream 5' elements, the stop codon, and downstream 3' elements are labelled above the nucleotide sequences. The conserved hexadecapeptide sequence [54] is highlighted and the amino acid residues indicated below the sequences. SNPs between the parental sequences are highlighted in red text and an asterisk (exons), with the one non-synonymous change surrounded by a box. Td-g = *Tragopogon dubius*-genomic; Td-c = *Tragopogon dubius*-cDNA; Tp-g = *Tragopogon pratensis*-genomic; Tp-c = *Tragopogon pratensis*-cDNA. FUE = far upstream element, NUE = near upstream element.

doi:10.1371/journal.pone.0144339.g001

fragment sizes 327 bp, 272 bp and 29 bp). For PCRs of both genomic and cDNA, a digestion reaction was set up in a total volume of 10 μl containing 1 μl of the PCR product, 1X buffer 4 (New England Biolabs, USA), 100 μg/ml Bovine Serum Albumin and 20 Units of *MseI* enzyme (New England Biolabs, USA). Reactions were incubated at 37°C for 3 hours as specified by the manufacturer. The digested products were run on a 2% agarose gel, stained with ethidium bromide and analyzed using a Gel Doc 2000 system (Bio-Rad, UK). After establishing the protocols for the diploid parents, *rbcS-1* was PCR-amplified from the naturally occurring and synthetic polyploids of *T. miscellus* and digested following the same protocols. We also included an artificial hybrid DNA or cDNA template, which contained an equal mixture of the two parental DNAs or cDNAs for genomic and cDNA CAPS, respectively. As a control to verify equal expression of parental homeologs for cDNA CAPS, actin was amplified (actinF: 5'-GGAGCAGAGAGATTCCGTTG-3', actinR: 5'-CTCTCTGGAGGAGCAACCA C-3') and digested with *BspHI* (S2 Fig) following the CAPS protocol above. The PCR conditions used to amplify actin were the same as those for *rbcS-1*.

5′ Genome walking and 3' RACE of *rbcS-1*

To obtain full-length *rbcS-1* sequence, we employed a 5'genome walking technique to amplify upstream unknown gene sequence (using a homemade kit following the GenomeWalker manual, Clontech Laboratories) [57]. Two outward-facing gene-specific primers were designed near the 5' end of the *T. dubius rbcS-1* sequence to act as reverse primers (GS1 and GS2, Table 1). Long and short oligos to form an adapter and adapter-specific primers (to act as forward primer) were designed as described by the GenomeWalker user manual (NA44 and NA45, Table 1). *Tragopogon dubius* genomic DNA was digested with three different blunt-cutting enzymes: *EcoRV*, *ScaI* and *DraI* (New England Biolabs) independently using 2.5 μg of genomic DNA, 80 Units of restriction enzyme and 10X buffer (New England Biolabs) in a total volume of 100 μl. Reactions were incubated at 37°C for 16–18 hours. These digestion reactions were cleaned by ethanol precipitation in the presence of 20 μg glycogen and 3M sodium acetate. Adapters were ligated to the cleaned, digested genomic DNA in a total volume of 8 μl containing 25 μM adapter, 10X ligation buffer, 3 Units of T4 DNA ligase (New England Biolabs) and 0.5 μg of purified DNA. Primary PCR was conducted in 50-μl total volume using 10 mM dNTPs, 10X PCR buffer (Takara Biotechnology, Japan), 10 μM adapter primer AP1 (Forward) and gene-specific primer GS1 (Reverse) (Table 1) and 1 Unit of Takara Ex Taq polymerase (Takara Biotechnology, Japan). Cycling conditions for the primary PCR were as follows: first 7 cycles at 94°C for 25 sec, 72°C for 3 min, then the remaining 32 cycles at 94°C for 25 sec, 67°C for 3 min, then final extension at 67°C for 7 min. Primary PCR products from the first round were diluted 1:50 in ddH_2O. In the secondary PCR, 10 μM nested or internal adapter primer AP2 (forward) and gene-specific primers GS2 (reverse) were used (Table 1), and 2 μl of diluted primary PCR product were used as template. The secondary PCR profile was as follows: 94°C for 25 sec, 72°C for 3 min for 5 cycles and 94°C for 25 sec, 67°C for 3 min for the next 20 cycles, then final extension at 67°C for 7 min. Secondary PCR products were separated on a 1% agarose gel, and products from each library were

cloned and sequenced using the protocols described above. The resulting sequences were aligned to the previously obtained partial *rbcS-1* sequence of *T. dubius*.

To obtain the 3' end of the *rbcS-1* gene, 3' RACE was used. Two gene-specific nested inverse primers were designed near the 3' end of the known *rbcS-1* gene sequence (Inv. Fwd 1 and Inv. Fwd 2, Table 1). First-strand cDNA from *T. dubius* was made using an oligo(dt) incorporating a 3' RACE-specific primer sequence at the 5' end (3' RACE oligodT adapter, Table 1). After synthesizing *T. dubius* cDNA, primary PCR for 3' RACE was conducted in a 25-μl total volume containing 5 μM gene-specific inverse primer (Inv. Fwd 2) as a forward primer and 5 μM 3' RACE adapter primer as a reverse primer, 10X PCR buffer, 10 mM dNTPs and 1 Unit Takara Ex *Taq* polymerase. The PCR profile was as follows: 95°C for 1 min, 53°C for 1 min, 72°C for 1 min for 5 cycles, followed by 44 cycles at 95°C for 1 min, 48°C for 1 min, 72°C for 1 min, with a final extension at 72°C for 7 min. This primary PCR product was diluted 100X and used as template for nested PCR. The nested PCR mix contained all of the above reagents, except 5 μM nested primers (Inv. Fwd 1 and 3' RACE adapter primer, Table 1) was used. Cycling conditions were the same as the 3' RACE primary PCR. Products from the nested PCR were cloned, sequenced and aligned with the previous *rbcS-1* gene sequence from *T. dubius*. Once the complete *rbcS-1* gene sequence for *T. dubius* was obtained, new primers were designed (*rbcS*-F1 and *rbcS*-R3, Table 1) for the amplification and sequencing of the complete *rbcS-1* gene from genomic DNA and cDNA of *T. pratensis*.

Prediction of *rbcS-1* gene structure

Gene structure of *rbcS-1* was predicted using Augustus (Version 2.6) [58] and GENSCAN [59]. These programs were used to confirm the transcription start site (TSS), exons, introns and other regulatory sequences as determined by cDNA sequencing of the complete *rbcS-1* gene. Plant Promoter Analysis Navigator (PlantPAN) [60] was used to identify promoter sequences of *rbcS-1*, putative transcription factor binding sites in the promoter region and conserved motifs in the promoter (S2 Table).

Homeolog-specific RT-PCR

Homeolog-specific RT-PCR was conducted to amplify each of the diploid parental homeologs of *rbcS-1* from cDNA of the *Tragopogon miscellus* polyploids. Homeolog-specific (HS) primers were based on SNPs identified between parental *rbcS-1* homeologs [61]. Homeolog specificity was assured by adding a mismatch one nucleotide away from the 3' end of each of the two forward HS primers (*T. dubius*: HS-dub-Snp3A, *T. pratensis*: HS-Pra-Snp3, Table 1). A common reverse primer was designed downstream of the polymorphic site (HS-R3, Table 1), corresponding to a highly conserved region in exon 3 (Fig 1). For the present experiment, forward primers were designed at the third SNP in exon 1 (corresponding to position 471 bp, Fig 1). PCR conditions were as follows: 95°C for 1 min, 60°C for 45 sec, 72°C for 1 min for 35 cycles with a final extension at 72°C for 10 min. PCR was conducted in a 25-μl total volume containing 10X PCR buffer, 10 mM dNTPs, 5 μM each primer, 0.5 Unit *Taq* polymerase (New England Biolabs, USA) and 15 ng/μl template (cDNA). The amount of template cDNA included in the PCR was quantified using a Nanodrop ND-1000 spectrophotometer (Thermo Fisher Scientific, USA) and normalized in the PCR reaction. Resulting PCR products were run on a 1.5% agarose gel and scored for presence/absence of parental homeologs or inspected to determine relative intensity of resulting bands. Different numbers of PCR cycles (30, 35, 40, and 45) were tested to determine the potential effect of amplification cycles. As no differences were observed among the different cycle numbers, the same PCR profile was used for all cDNA amplifications. Artificial hybrid cDNA, which was a 50:50 mix of *T. dubius* and *T. pratensis* cDNA, was again included as a control for equal amplification of the parental homeologs.

Results

rbcS gene family

Two *rbcS* gene copies, *rbcS-1* and *rbcS-2*, were identified from the *Tragopogon dubius* EST database. These two *rbcS* genes are fairly divergent from each other, with several SNPs, insertions and deletions in the genic regions and even more variability at the 3' UTRs (S1 Fig). Of these two *rbcS* genes, *rbcS-1* was determined to be a functional copy, while the second *rbcS* gene, *rbcS-2*, was considered a pseudogene and a truncated copy as it had premature stop codons compared to the *rbcS-1* amino acid alignment. Several attempts at 5'genome walking experiments yielded non-*rbcS* genomic sequences (e.g., plastid *atpB* sequence) at its flanking ends where conserved sequence would have been expected. We found a similar scenario with *T. pratensis*. Hence, we focused on *rbcS-1* to examine potential cytonuclear coordination in *T. miscellus*. Sequences of *rbcS-1* sequences of *T. dubius* and *T. pratensis* were deposited in GenBank (accession numbers: KT879189, KT879190, respectively).

The total length of the *rbcS-1* sequence with coding and non-coding regions, including upstream promoter elements and downstream terminator signals, was 1212 bp in *Tragopogon dubius* and 1219 bp in *T. pratensis*. Fig 1 shows the full-length genomic sequences of these two diploid species with the conserved 5' upstream elements, canonical hexadecapeptide sequence [54] and SNPs between them highlighted. No intraspecific variation in genomic sequences of *rbcS-1* was detected among individuals of either *Tragopogon dubius* or *T. pratensis*.

Divergence of *rbcS-1* and *rbcL* between the diploids and their pattern of inheritance and retention in *T. miscellus*

Comparative sequence analysis of *rbcS-1* in *T. dubius* and *T. pratensis* revealed seven SNPs in the exons and a 1-bp indel in the 3' UTR between polyadenylation signals at 1114 bp (Fig 1). Six of these SNPs were synonymous substitutions, with the second SNP at 424 bp a non-synonymous change resulting in a threonine in *T. pratensis* and a serine in *T. dubius*. Non-coding regions (upstream promoter regions and introns) were also found to contain multiple SNPs and indels (Fig 1) between the diploids. Analysis of the predicted protein structure of the *rbcS-1* sequence using the protein homology/analogy recognition engine Phyre^2 V 2.0 [62] revealed that the non-synonymous SNP resides in an alpha-helix and does not cause any difference in predicted protein structure between *rbcS-1* parental homeologs. Genomic *rbcL* (1415 bp) sequences from both diploid parents were compared, and only one SNP was discovered at 703 bp, resulting in a synonymous substitution (Genbank accessions: KT897489, KT897491).

To determine the pattern of retention of these subunits, genomic copies of both *rbcS-1* and *rbcL* were analyzed from 25 synthetic polyploid individuals (representing five independently generated lineages) and 79 naturally occurring polyploids from 10 populations of *T. miscellus*. In the case of *rbcL*, all synthetic and naturally occurring polyploids had the maternally derived sequence (i.e., *T. pratensis* for the short-liguled form and *T. dubius* for the long-liguled form; *T. miscellus* Genbank accessions: KT897488, KT897490). For *rbcS-1*, all synthetic polyploids and 69 of the naturally occurring polyploids had both *T. dubius* and *T. pratensis rbcS-1* homeologs, as determined by additivity of the genomic CAPS analysis (Fig 2A). Inspection of the chromatograms resulting from directly sequenced *rbcS-1* products from these same *T. miscellus* individuals also revealed additivity of peaks at SNPs between the parents. Ten polyploid individuals from six natural populations [Spangle (2), Garfield (1), Albion (3), Moscow (2), Pullman (1) and Troy (1)] had only one homeolog present in the genomic DNA. Of nine short-liguled individuals, six had the maternally derived *rbcS-1* homeolog (*T. pratensis*), and three had the paternally derived copy (*T. dubius*, Fig 2A, Table 2, S1 Table). One long-liguled

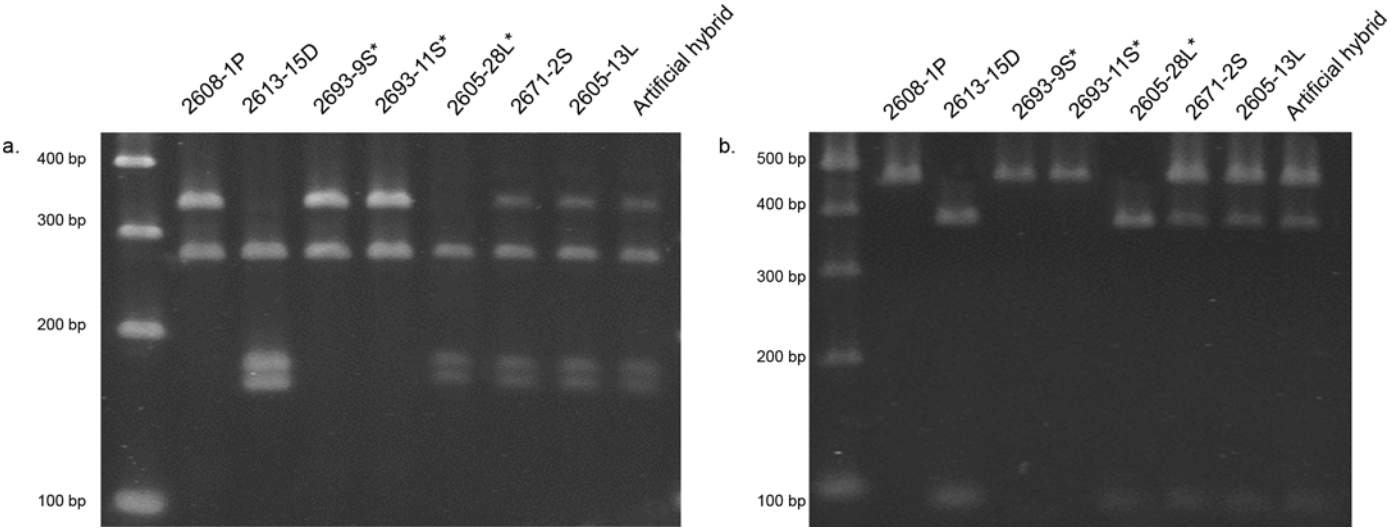

Fig 2. Genomic (a) and cDNA (b) cleaved amplified polymorphic sequence (CAPS) results for representative samples of naturally occurring *Tragopogon miscellus* polyploids and the diploid parents, *T. dubius* (D) and *T. pratensis* (P). *T. pratensis* is the maternal parent of the short-liguled (S) individuals, and *T. dubius* is the maternal parent of the long-liguled (L) individuals. An asterisk (*) indicates homeolog loss in *T. miscellus*. The artificial hybrid contained equal mixture of *T. dubius* and *T. pratensis* genomic DNA (a) or cDNA (b). Population codes are detailed in S1 Table.

doi:10.1371/journal.pone.0144339.g002

individual from Pullman retained the *T. dubius* (maternal) genomic homeolog only (Fig 2A, Table 2).

Expression of *rbcS-1* homeologs in *T. miscellus* polyploids

Relative expression of parental *rbcS-1* homeologs was determined by cDNA CAPS (Fig 2B) and HS-RT-PCR (Fig 3) analyses. Because the cDNA CAPS results did not appear to show any

Table 2. Naturally occurring individuals of *Tragopogon miscellus* that showed bias in the retention and/or expression of parental *rbcS-1* homeologs.

Population	Maternal parent	Lineage	Retention of *rbcS* homeologs	Expression of *rbcS* homeologs
Spangle	*T. pratensis*	2693–7	Both	*T. pratensis* > *T. dubius*
Spangle	*T. pratensis*	2693–9	*T. pratensis* only	*T. pratensis* only
Spangle	*T. pratensis*	2693–11	*T. pratensis* only	*T. pratensis* only
Oakesdale	*T. pratensis*	2671–2	Both	*T. pratensis* > *T. dubius*
Oakesdale	*T. pratensis*	2671–11	Both	*T. pratensis* > *T. dubius*
Garfield	*T. pratensis*	2688–8	*T. pratensis* only	*T. pratensis* only
Garfield	*T. pratensis*	2688–12	Both	*T. pratensis* > *T. dubius*
Moscow	*T. pratensis*	2604–17	*T. pratensis* only	*T. pratensis* only
Moscow	*T. pratensis*	2604–22	*T. pratensis* only	*T. pratensis* only
Moscow	*T. pratensis*	2604–43	Both	*T. pratensis* > *T. dubius*
Albion	*T. pratensis*	2625–3	*T. dubius* only	-
Albion	*T. pratensis*	2625–6	*T. dubius* only	-
Albion	*T. pratensis*	2625–8	*T. dubius* only	-
Troy	*T. pratensis*	2682–5	*T. pratensis* only	-
Pullman	*T. dubius*	2605–9	Both	*T. dubius* > *T. pratensis*
Pullman	*T. dubius*	2605–28	*T. dubius* only	*T. dubius* only
Pullman	*T. dubius*	2605–46	Both	*T. dubius* > *T. pratensis*

A dash (-) indicates that material was not available to study a particular individual for both retention (genomic DNA) and expression (cDNA).

doi:10.1371/journal.pone.0144339.t002

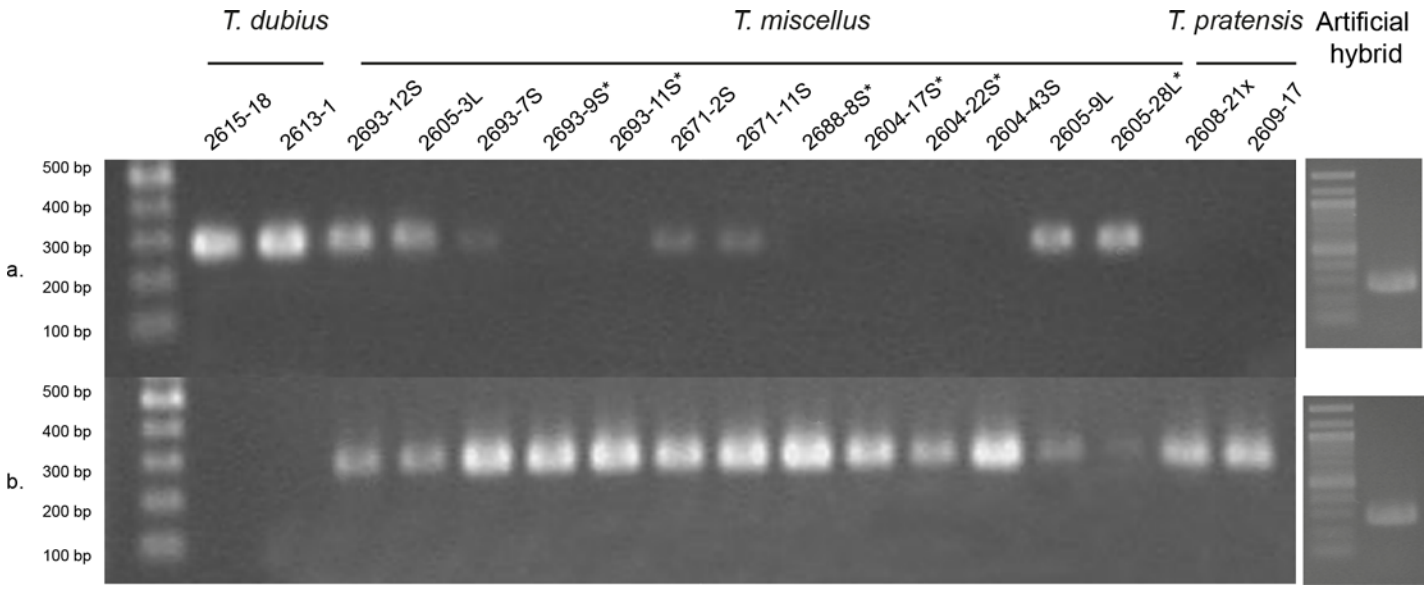

Fig 3. Homeolog-specific HS-RT-PCR of *rbcS-1* for representative individuals of *Tragopogon miscellus* and diploid progenitors *T. dubius* and *T. pratensis*. RT-PCR results using *T. dubius*-specific primers (a) and *T. pratensis*-specific primers (b). *T. pratensis* is the maternal parent of the short-liguled (S) individuals, and *T. dubius* is the maternal parent of the long-liguled (L) individuals. An asterisk (*) indicates homeolog loss in *T. miscellus*. The artificial hybrid contained equal mixture of *T. dubius* and *T. pratensis* cDNA. Population codes are detailed in S1 Table.

doi:10.1371/journal.pone.0144339.g003

appreciable differences in homeolog expression, HS-RT-PCR was employed as a potentially more sensitive method to detect differential homeolog expression. For the six synthetic polyploid individuals examined, all showed equal expression of the parental homeologs (S1 Table). Of the 31 naturally occurring polyploid individuals examined, five showed deviation from additive expression (Table 2, Figs 2B and 3), as determined by comparison of the relative intensity or presence/absence of parental homeologs in the polyploids, using the positive controls as a baseline for those comparisons. In all cases where both genomic copies were detected, the maternally derived *rbcS-1* homeolog was expressed at a greater level (Fig 3). The five individuals showing differential expression were from different populations and represented one long-liguled individual from Pullman and four short-liguled individuals [Moscow (1), Oakesdale (2), and Spangle (1)]. All *T. miscellus* individuals that showed biased expression of the *rbcS-1* homeologs were found to have equal expression of *T. dubius* and *T. pratensis* actin homeologs (S2 Fig).

Discussion

Characterization of *rbcS-1* in *Tragopogon* diploid species

In angiosperms, the rbcS small subunit is fairly divergent among species and is often encoded by a multigene nuclear family [63–67], compared to the plastid *rbcL*, which is highly conserved and present in single copy [66]. Indeed, only one SNP distinguished the *Tragopogon* progenitor *rbcL* copies, and this resulted in a synonymous substitution. As is the case in most other eudicots [30, 63], the *rbcS-1* gene in *Tragopogon* consists of three exons separated by two short introns. The second copy found in *T. dubius* and *T. pratensis* (*rbcS-2*) may represent a pseudogene on its way to being lost from the genome. In most other angiosperms, the *rbcS* gene family ranges in size from four (*Arabidopsis*) to more than 22 (wheat) copies [64, 67, 68]. Generally, only one or two members of the *rbcS* gene family are strongly expressed in the angiosperms surveyed to date, and these genes contribute more than half of the total *rbcS* transcripts [69,

70]. Indeed, concerted evolution of the *rbcS* gene family is probably a common phenomenon even in diploid taxa [30]. In other members of Asteraceae, *Lactuca sativa* (tribe Cichorieae) has six *rbcS* genes [71], and *Flaveria* species (Heliantheae) contain from 5–16 genes [72], while *Helianthus* (Heliantheae) [73] and *Chrysanthemum* (Anthemideae) [69] each has only one *rbcS* gene. Thus, *rbcS* may be diverse in copy number even within the same plant family or tribe (*Tragopogon* is a member of the Cichorieae), perhaps due to general processes of gene loss, genome downsizing, concerted evolution or a mere lack of expansion of the gene family [74, 75].

Interspecific *rbcS-1* sequence variation was low between the parental diploids *Tragopogon dubius* and *T. pratensis* (2.5% sequence divergence, 0.5% amino acid divergence), compared to other genera (*Gossypium* [31], *Arabidopsis* [64], *Triticum* [68]). Only one non-synonymous substitution was detected between *T. dubius* and *T. pratensis rbcS-1* homeologs; this SNP resided in the α-helix of the predicted protein and did not result in a change in protein structure or folding. It is not obvious if this change has an influence on the seemingly maternal bias toward homeolog retention and expression in the *T. miscellus* polyploids.

Genomic loss and expression of *rbcS-1* homeologs biased towards the maternal parent in *T. miscellus* polyploids

From an evolutionary perspective, the dynamic nature of polyploid genomes is well known [8, 76, 77]. Homeolog loss is one genetic modification commonly observed following genome duplication in a diverse array of polyploid species (*Brassica* [11], *Triticum* [14], *Tragopogon* [44], *Gossypium* [78], *Arabidopsis* [79]). The results presented here are generally consistent with previous findings of preferential retention and expression of maternal homeologs in *Tragopogon* [40, 44, 45, 80]. In polyploids, homeolog losses may be associated with dosage compensation to efficiently maintain gene regulatory mechanisms [81]. In the case of cytonuclear coordination involving multi-subunit complexes, like Rubisco, loss of the paternal homeolog and retention of the maternal copy may facilitate the regulatory coordination between the maternal and paternal genomes. However, in *Tragopogon*, this coordination is not immediate upon polyploid formation as the synthetics and most of the naturally occurring polyploids still retain and express both parental homeologs. Compared to other polyploid systems that are much older [30, 31], we also did not find any evidence for unique mutations (autapomorphies) in the natural or synthetic *Tragopogon miscellus rbcS* sequences. Examination of other cytonuclear complexes would lend insight to the potential to maintain genomic balance between nuclear homeologs and their cytoplasmic counterparts.

Changes in duplicate gene expression are another consequence of allopolyploidization [1, 8, 82, 83], which may involve biased expression of the parental homeologs in the polyploids. This bias may be balanced, with an equal number of genes showing bias towards each parent, or unbalanced, with more genes displaying bias towards one parent [28, 84–87]. Previous studies on *Tragopogon* identified alterations in expression of homeologous loci (i.e., *T. dubius* loci silenced more often than *T. pratensis*) [1, 44, 88]. In this study, expression of parental *rbcS-1* homeologs was biased toward the maternal parent, although again very few individuals showed this pattern and it was not immediately upon allopolyploid formation.

The successful establishment of F_1 hybrids and allopolyploids requires coordination between the maternally inherited cytoplasmic (plastid and mitochondrial) and the biparentally inherited nuclear genomes to facilitate genomic stability [22, 89]. Cytoplasmic factors, including a variety of nucleo-cytoplasmic co-evolutionary pathways, have been considered responsible for post-zygotic hybrid incompatibilities and therefore a driver of plant speciation [23]. Here we show that this coordination may be a slower process and does not occur immediately

upon formation in *Tragopogon*, however, given that the naturally occurring *Tragopogon miscellus* populations are less than 80 years old (~40 generations as they are biennials), sorting out potential cytonuclear incompatibilities may only take a few generations to begin. Examination of the synthetic lineages over successive generations would lend valuable insight as to when these changes start to occur.

The biased retention and expression of maternal *rbcS-1* homeologs in individuals from different populations indicates repeatability of this evolutionary trajectory because each population of *T. miscellus* represents an independent formation [33, 36, 38, 90]. However, within a population, the observed homeolog losses may result from the same historical event; thus, our estimates of absolute losses may be lower (six rather than ten). Although the majority of the individuals showed maternal bias, three individuals from the Albion population were an exception. These short-liguled individuals retained the paternal (*T. dubius*) *rbcS-1* genomic homeolog, instead of the *T. pratensis* copy. Unfortunately, fresh material was not available to study *rbcS-1* expression in individuals from this population, so we do not know if the paternal bias is restricted to genome loss or extends to homeolog expression as well for individuals that retained both homeologs. In a previous study of homeolog loss in *T. miscellus* [45], this population showed a greater number of homeolog losses (individual 2625–3 in particular) than all other populations. Although in general there seems to be a recurrent pattern toward maternal bias, in some populations different *rbcS/rbcL* parental combinations might be beneficial to facilitate cytonuclear interactions.

Given that the predicted protein structure of both parental *rbcS-1* homeologs is the same and the *rbcL* progenitor copies only differ by one synonymous SNP, exactly what has driven differential expression of the maternal copy of *rbcS-1* in the naturally occurring polyploids is not yet understood. There are several possible explanations for the expression biases observed. First, the polymorphisms observed between *rbcS-1* homeologs in the promoter region (e.g., one SNP was found eight nucleotides away from the transcription start site) might be responsible for differential regulation of *rbcS-1* homeologs, and later, their interaction with the *rbcL*-encoded subunit. Dean *et al.* [63] found that specific *rbcS* copies in *Petunia* (Solanaceae) contained 'enhancer-like' elements in the promoter region that resulted in quantitative differences in expression levels, even when there was a high degree of similarity in coding sequence among other copies. This region, termed box II, was also identified in other solanaceous genera (tomato, *Solanum*, and tobacco, *Nicotiana*), and *rbcS* copies with this motif were expressed at a greater level than were other copies. *Tragopogon* also contains this enhancer-like motif, but no SNPs between the parents were identified in this region. Perhaps the other polymorphisms in the promoter region contribute to the expression differences observed here.

A second explanation for the expression bias is that the one non-synonymous change in exon I (threonine in *T. pratensis* and serine in *T. dubius*) may result in differential selection on the *rbcS-1* copies under some conditions. Further research involving protein-protein interactions between rbcS/rbcL subunits in *T. miscellus* would be helpful to clarify the complexities of these cytonuclear interactions.

Supporting Information

S1 Fig. Alignment of *rbcS* cDNA sequences and protein translations for *Tragopogon dubius*, *T. pratensis* and selected Asteraceae. The conserved hexadecapeptide motif (YYDGRYWTMWKLPMFG) is indicated in red text.
(PDF)

S2 Fig. cDNA-CAPS of actin to verify equal expression of parental copies in *T. miscellus* polyploids. A.H. stands for artificial hybrid which was a 1:1 mixture of *T. dubius* and *T.*

pratensis cDNA.
(PDF)

S1 Table. Data summary for all *Tragopogon miscellus* allotetraploids and diploid parents (*T. dubius* and *T. pratensis*) examined. Data are summarized from genomic DNA and cDNA CAPS and homeolog-specific RT-PCR. Note: Letters "D" and "P" correspond to the diploid parents *T. dubius* and *T. pratensis*, respectively. A 'D' or a 'P' indicates that only one parental homeolog was detected in genomic DNA or expressed. P>D indicates that the *T. pratensis* homeolog showed higher relative expression than the *T. dubius rbcS-1* homeolog in the *T. miscellus* individual and vice versa for D>P.
(PDF)

S2 Table. Transcription factor binding sites identified in the *Tragopogon rbcS-1* promoter region.
(PDF)

Acknowledgments

We thank Muhammad Faisal for assistance in predicting promoter and terminator elements, Afsana Islam for her guidance in genome walking experiments, and Prashant Joshi for technical assistance. This work was supported in part by a Massey University Research Fund grant to JT. The Higher Education Commission (HEC) of Pakistan is gratefully acknowledged for a Ph. D. scholarship to TS.

Author Contributions

Conceived and designed the experiments: TS JT. Performed the experiments: TS. Analyzed the data: TS VS DS PS JT. Contributed reagents/materials/analysis tools: DS PS JT. Wrote the paper: TS VS DS PS JT.

References

1. Buggs RJA, Chamala S, Wu W, Gao L, May GD, Schnable PS, et al. Characterization of duplicate gene evolution in the recent natural allopolyploid *Tragopogon miscellus* by next-generation sequencing and Sequenom iPLEX MassARRAY genotyping. Mol Ecol 2010; 19: 132–146. doi: 10.1111/j.1365-294X.2009.04469.x
2. Te Beest M, Le Roux JJ, Richardson DM, Brysting AK, Suda J, Kubesova M, et al. The more the better? The role of polyploidy in facilitating plant invasions. Ann bot 2012; 109: 19–45. doi: 10.1093/aob/mcr277 PMID: 22040744
3. Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng CF, et al. Polyploidy and angiosperm diversification. Am J Bot 2009; 96: 336–348. doi: 10.3732/ajb.0800079 PMID: 21628192
4. Yang XH, Ye CY, Cheng ZM, Tschaplinski TJ, Wullschleger SD, Yin WL, et al. Genomic aspects of research involving polyploid plants. Plant Cell Tiss Organ 2011; 104: 387–397. doi: 10.1007/s11240-010-9826-1
5. Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, et al. Ancestral polyploidy in seed plants and angiosperms. Nature 2011; 473: 97–100. Epub 100. doi: 10.1038/nature09916 PMID: 21478875
6. Yoo MJ, Szadkowski E, Wendel JF. Homoeolog expression bias and expression level dominance in allopolyploid cotton. Heredity 2013; 110: 171–180. doi: 10.1038/hdy.2012.94 PMID: 23169565
7. Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC. Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. Plant Cell 2007; 19: 3403–3417. PMID: 18024568
8. Chen K, Rajewsky N. The evolution of gene regulation by transcription factors and microRNAs. Nature Reviews Genetics 2007; 8: 93–103. PMID: 17230196

9. Comai L, Tyagi AP, Winter K, Holmes-Davis R, Reynolds SH, Stevens Y, et al. Phenotypic instability and rapid gene silencing in newly formed *Arabidopsis* allotetraploids. Plant Cell 2000; 12: 1551–1567. PMID: 11006331
10. Ma XF, Gustafson JP. Allopolyploidization-accommodated genomic sequence changes in Triticale. Ann Bot 2008; 101: 825–832. doi: 10.1093/aob/mcm331 PMID: 18252766
11. Xiong Z, Gaeta RT, Pires JC. Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. PNAS 2011; 108: 7908–7913. doi: 10.1073/pnas.1014138108 PMID: 21512129
12. Hegarty MJ, Barker GL, Wilson ID, Abbott RJ, Edwards KJ, Hiscock SJ. Transcriptome shock after interspecific hybridization in *Senecio* is ameliorated by genome duplication. Curr Biol 2006; 16: 1652–1659. http://dx.doi.org/10.1016/j.cub.2006.06.071 PMID: 16920628
13. Chester M, Gallagher JP, Symonds VV, da Silva AVC, Mavrodiev EV, Leitch AR, et al. Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon miscellus* (Asteraceae). PNAS 2012; 109: 1176–1181. doi: 10.1073/pnas.1112041109 PMID: 22228301
14. Feldman M, Levy A, Chalhoub B, Kashkush K. Genomic plasticity in polyploid wheat. In: Soltis PS, Soltis DE, editors. Polyploidy and genome evolution. Berlin Heidelberg: Springer 2012. p. 109–135.
15. Prince VE, Pickett FB. Splitting pairs: the diverging fates of duplicated genes. Nat Rev Genet 2002; 3: 827–837. PMID: 12415313
16. Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, et al. Evolutionary genetics of genome merger and doubling in plants. Annu Rev Genet 2008; 42: 443–461. doi: 10.1146/annurev.genet.42.110807.091524 PMID: 18983261
17. Conant GC, Wolfe KH. Turning a hobby into a job: How duplicated genes find new functions. Nat Rev Genet 2008; 9: 938–950. doi: 10.1038/nrg2482 PMID: 19015656
18. Edger PP, Pires JC. Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. Chromosome Res 2009; 17: 699–717. doi: 10.1007/s10577-009-9055-9 PMID: 19802709
19. Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. Science 2000; 290: 1151–1155. PMID: 11073452
20. Blanc G, Wolfe KH. Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. Plant Cell 2004; 16: 1679–1691. PMID: 15208398
21. Roulin A, Auer PL, Libault M, Schlueter J, Farmer A, May G, et al. The fate of duplicated genes in a polyploid plant genome. Plant J 2013; 73: 143–153. doi: 10.1111/tpj.12026 PMID: 22974547
22. Fishman L, Willis JH. A cytonuclear incompatibility causes anther sterility in *Mimulus* hybrids. Evolution 2006; 60: 1372–1381. doi: 10.1554/05-708.1 PMID: 16929654
23. Levin DA. The cytoplasmic factor in plant speciation. Syst Bot 2003; 28: 5–11. doi: 10.2307/3093933
24. Sloan DB. Using plants to elucidate the mechanisms of cytonuclear co-evolution. New Phytol 2015; 205: 1040–1046. PMID: 25729802
25. Oehlkers F. Cytoplasmic inheritance in the genus *Streptocarpus lindley*. Adv Genet 1964; 12: 329–370. doi: 10.1016/s0065-2660(08)60418-6
26. Grant V. The genetic structure of races and species in *Gilia*. Adv Genet 1956; 8: 55–87. doi: 10.1016/s0065-2660(08)60499-x
27. Song KM, Lu P, Tang KL, Osborn TC. Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. PNAS 1995; 92: 7719–7723. PMID: 7644483
28. Adams KL, Cronn R, Percifield R, Wendel JF. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. PNAS 2003; 100: 4649–4654. PMID: 12665616
29. Wolf JB. Cytonuclear interactions can favor the evolution of genomic imprinting. Evolution 2009; 63: 1364–1371. PMID: 19425202
30. Gong L, Olson M, Wendel JF. Cytonuclear evolution of Rubisco in four allopolyploid lineages. Mol Biol Evol 2014; 31: 2624–2636. doi: 10.1093/molbev/msu207 PMID: 25015644
31. Gong L, Salmon A, Yoo M-J, Grupp KK, Wang Z, Paterson AH, et al. The cytonuclear dimension of allopolyploid evolution: An example from cotton using rubisco. Mol Biol Evol 2012; 29: 3023–3036. doi: 10.1093/molbev/mss110 PMID: 22490824
32. Ownbey M. Natural hybridization and amphiploidy in the genus *Tragopogon*. Am J Bot 1950; 37: 487–499.
33. Soltis DE, Soltis PS. Polyploidy: recurrent formation and genome evolution. Trends in Ecol Evol 1999; 14: 348–352.

34. Ownbey M, McCollum GD. Cytoplasmic inheritance and reciprocal amphiploidy in *Tragopogon*. Am J Bot 1953; 40: 788–796.

35. Soltis DE, Soltis PS. The dynamic nature of polyploid genomes. PNAS 1995; 92: 8089–8091. PMID: 7667249

36. Soltis DE, Soltis PS, Pires JC, Kovarik A, Tate JA, Mavrodiev E. Recent and recurrent polyploidy in *Tragopogon* (Asteraceae): cytogenetic, genomic and genetic comparisons. Biol J Linn Soc 2004; 82: 485–501.

37. Soltis PS, Soltis DE. Multiple origins of the allotetraploid *Tragopogon mirus* (Compositae)—rDNA evidence. Syst Bot 1991; 16: 407–413. doi: 10.2307/2419333

38. Symonds VV, Soltis PS, Soltis DE. Dynamics of polyploid formation in *Tragopogon* (Asteraceae): recurrent formation, gene flow, and population structure. Evolution 2010; 64: 1984–2003. doi: 10.1111/j.1558-5646.2010.00978-x PMID: 20199558

39. Buggs RJA, Chamala S, Wu W, Tate JA, Schnable PS, Soltis DE, et al. Rapid, repeated, and clustered loss of duplicate genes in allopolyploid plant populations of independent origin. Curr Biol 2012; 22: 248–252. doi: http://dx.doi.org/10.1016/j.cub.2011.12.027 PMID: 22264605

40. Buggs RJA, Doust AN, Tate JA, Koh J, Soltis K, Feltus FA, et al. Gene loss and silencing in *Tragopogon miscellus* (Asteraceae): comparison of natural and synthetic allotetraploids. Heredity 2009; 103: 73–81. doi: 10.1038/hdy.2009.24 PMID: 19277058

41. Buggs RJA, Renny-Byfield S, Chester M, Jordon-Thaden IE, Viccini LF, Chamala S, et al. Next-generation sequencing and genome evolution in allopolyploids. Am J Bot 2012; 99: 372–382. doi: 10.3732/ajb.1100395 PMID: 22268220

42. Buggs RJA, Zhang LJ, Miles N, Tate JA, Gao L, Wei W, et al. Transcriptomic shock generates evolutionary novelty in a newly formed, natural allopolyploid plant. Curr Biol 2011; 21: 551–556. doi: 10.1016/j.cub.2011.02.016 PMID: 21419627

43. Koh J, Chen S, Zhu N, Yu F, Soltis PS, Soltis DE. Comparative proteomics of the recently and recurrently formed natural allopolyploid *Tragopogon mirus* (Asteraceae) and its parents. New Phyto 2012; 196: 292–305. doi: 10.1111/j.1469-8137.2012.04251.x

44. Koh J, Soltis PS, Soltis DE. Homeolog loss and expression changes in natural populations of the recently and repeatedly formed allotetraploid *Tragopogon mirus* (Asteraceae). BMC Genomics 2010; 11: 97. doi: 10.1186/1471-2164-11-97 PMID: 20141639

45. Tate JA, Joshi P, Soltis KA, Soltis PS, Soltis DE. On the road to diploidization? Homoeolog loss in independently formed populations of the allopolyploid *Tragopogon miscellus* (Asteraceae). BMC Plant Biol 2009; 9: 80. doi: 10.1186/1471-2229-9-80 PMID: 19558696

46. Tate JA, Ni Z, Scheen A-C, Koh J, Gilbert CA, Lefkowitz D, et al. Evolution and expression of homeologous loci in *Tragopogon miscellus* (Asteraceae), a recent and reciprocally formed allopolyploid. Genetics 2006; 173: 1599–1611. doi: 10.1534/genetics.106.057646 PMID: 16648586

47. Lim KY, Soltis DE, Soltis PS, Tate J, Matyasek R, Srubarova H, et al. Rapid chromosome evolution in recently formed polyploids in *Tragopogon* (Asteraceae). PLoS ONE 2008; 3: e3353.

48. Tate JA, Symonds VV, Doust AN, Buggs RJA, Mavrodiev E, Majure LC, et al. Synthetic polyploids of *Tragopogon miscellus* and *T. mirus* (Asteraceae): 60 Years after Ownbey's discovery. Am J Bot 2009; 96: 979–988. doi: 10.3732/ajb.0800299 PMID: 21628250

49. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 1987; 19: 11–15.

50. Hillis DM, Moritz C, Mable BK, editors. Molecular systematics. Sunderland, MA, USA: Sinauer Associates, Inc; 1996.

51. Olmstead RG, Michaels HJ, Scott KM, Palmer JD. Monophyly of the Asteridae and identification of their major lineages inferred from DNA sequences of *rbcL*. Ann MO Bot Gard 1992; 79: 249–265. doi: 10.2307/2399768

52. Panero JL, Crozier BS. Primers for PCR amplification of Asteraceae chloroplast DNA. Lundellia 2003; 6: 1–9.

53. Rozen S, Skaletsky HJ. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S, editors. Bioinformatics Methods and Protocols: Methods in Molecular Biology Humana Press, Totowa, NJ, USA2002. p. 365–386. PMID: 10547847

54. Dean C, Elzen PVd, Tamaki S, Black M, Dunsmuir P, Bedbrook J. Molecular characterization of rbcS multigene family in *Petunia*. Mol gen Genet 1987; 206: 465–74.

55. Neff MM, Turk E, Kalishman M. Web-based primer design for single nucleotide polymorphism analysis. Trends Genet 2002; 18: 613–615. doi: 10.1016/s0168-9525(02)02820-2 PMID: 12446140

56. Vincze T, Posfai J, Roberts RJ. NEBcutter: a program to cleave DNA with restriction enzymes. Nucleic Acids Res 2003; 31: 3688–3691. doi: 10.1093/nar/gkg526 PMID: 12824395

57. Siebert PD, Chenchik A, Kellogg DE, Lukyanov KA, Lukyanov SA. An improved PCR method for walking in uncloned genomic DNA. Nucleic Acids Res 1995; 23: 1087–1088. doi: 10.1093/nar/23.6.1087 PMID: 7731798

58. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics 2008; 24: 637–644. doi: 10.1093/bioinformatics/btn013 PMID: 18218656

59. Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. J Mol Bio 1997; 268: 78–94.

60. Chang WC, Lee TY, Huang HD, Huang HY, Pan RL. PlantPAN: Plant promoter analysis navigator, for identifying combinatorial cis-regulatory elements with distance constraint in plant gene groups. BMC Genomics 2008; 9: 561. doi: 56110.1186/1471-2164-9-561 doi: 10.1186/1471-2164-9-561 PMID: 19036138

61. Li BH, Kadura I, Fu DJ, Watson DE. Genotyping with TaqMAMA. Genomics 2004; 83: 311–320. doi: 10.1016/j.ygeno.2003.08.005 PMID: 14706460

62. Kelley LA, Sternberg MJE. Protein structure prediction on the web: a case study using the *Phyre* server. Nature Protocols 2009; 4: 363–371. doi: 10.1038/nprot.2009.2 PMID: 19247286

63. Dean C, Pichersky E, Dunsmuir P. Structure, evolution and regulation of *rbcS* genes in higher plants. Ann Rev Plant Phys 1989; 40: 415–439. doi: 10.1146/annurev.pp.40.060189.002215

64. Krebbers E, Seurinck J, Herdies L, Cashmore A, Timko M. Four genes in two diverged subfamilies encode the ribulose-1,5-bisphosphate carboxylase small subunit polypeptides of *Arabidopsis thaliana*. Plant Mol Bio 1988; 11: 745–759. doi: 10.1007/bf00019515

65. Miziorko HM, Lorimer GH. Ribulose-1,5-bisphosphate carboxylase-oxygenase. Annu Rev Biochem 1983; 52: 507–535. doi: 10.1146/annurev.bi.52.070183.002451 PMID: 6351728

66. Spreitzer RJ, Salvucci ME. Rubisco: Structure, regulatory interactions, and possibilities for a better enzyme. Ann Rev Plant Biol 2002; 53: 449–475. doi: 10.1146/annurev.arplant.53.100301.135233

67. Spreitzer RJ. Role of the small subunit in ribulose-1,5-bisphosphate carboxylase/oxygenase. Arch Biochem Biophys 2003; 414: 141–149. doi: 10.1016/s0003-9861(03)00171-1 PMID: 12781765

68. Sasanuma T. Characterization of the *rbcS* multigene family in wheat: subfamily classification, determination of chromosomal location and evolutionary analysis. Mol Genet Genom 2001; 265: 161–171. doi: 10.1007/s004380000404

69. Outchkourov NS, Peters J, Jong J, Rademakers W, Jongsma MA. The promoter-terminator of chrysanthemum *rbcS1* directs very high expression levels in plants. Planta 2003; 216: 1003–1012. doi: 10.1007/s00425-002-0953-8 PMID: 12687368

70. Izumi M, Tsunoda H, Suzuki Y, Makino A, Ishida H. *RBCS1A* and *RBCS3B*, two major members within the *Arabidopsis RbcS* multigene family, function to yield sufficient rubisco content for leaf photosynthetic capacity. J Exp Bot 2012; 63: 2159–2170. doi: 10.1093/jxb/err434 PMID: 22223809

71. Goumenaki E, Taybi T, Borland A, Barnes J. Mechanisms underlying the impacts of ozone on photosynthetic performance. Environ Exp Bot 2010; 69: 259–266. doi: http://dx.doi.org/10.1016/j.envexpbot.2010.04.011

72. Kapralov MV, Kubien DS, Andersson I, Filatov DA. Changes in rubisco kinetics during the evolution of C4 photosynthesis in *Flaveria* (Asteraceae) are associated with positive selection on genes encoding the enzyme. Mol Bio Evol 2011; 28: 1491–1503. doi: 10.1093/molbev/msq335

73. Waksman G, Lebrun M, Freyssinet G. Nucleotide-sequence of a gene encoding sunflower ribulose-1,5-bisphosphate carboxylase oxygenase small subunit (*rbcS*). Nucleic Acids Res 1987; 15: 7181–7181. doi: 10.1093/nar/15.17.7181 PMID: 3658679

74. Wang X, Shi X, Hao B, Ge S, Luo J. Duplication and DNA segmental loss in the rice genome: implications for diploidization. New Phytol 2005; 165: 937–946. doi: 10.1111/j.1469-8137.2004.01293.x PMID: 15720704

75. Wolfe KH. Yesterday's polyploids and the mystery of diploidization. Nat Rev Genet 2001; 2: 333–341. PMID: 11331899

76. Wendel JF. Genome evolution in polyploids. Plant Mol Biol 2000; 42: 225–249. PMID: 10688139

77. Soltis PS, Soltis DE. The role of genetic and genomic attributes in the success of polyploids. PNAS 2000; 97: 7051–7057. doi: 10.1073/pnas.97.13.7051 PMID: 10860970

78. Wendel JF, Flagel LE, Adams KL. Jeans, genes, and genomes: cotton as a model for studying polyploidy. In: Soltis PS, Soltis DE, editors. Polyploidy and Genome Evolution. New York: Springer; 2012.

79. Matsushita SC, Tyagi AP, Thornton GM, Pires JC, Madlung A. Allopolyploidization lays the foundation for evolution of distinct populations: evidence from analysis of synthetic *Arabidopsis* allohexaploids. Genetics 2012; 191: 535. doi: 10.1534/genetics.112.139295 PMID: 22426881

80. Malinska H, Tate JA, Matyasek R, Leitch AR, Soltis DE, Soltis PS, et al. Similar patterns of rDNA evolution in synthetic and recently formed natural populations of *Tragopogon (*Asteraceae) allotetraploids. BMC Evol Bio 2010; 10: 291. doi: 291 doi: 10.1186/1471-2148-10-291

81. Birchler JA, Veitia RA. The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. New Phyto 2010; 186: 54–62. doi: 10.1111/j.1469-8137.2009.03087.x

82. Flagel LE, Wendel JF. Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation. New Phyto 2010; 186: 184–193. doi: 10.1111/j.1469-8137.2009.03107.x

83. Bottley A, Xia GM, Koebner RMD. Homoeologous gene silencing in hexaploid wheat. Plant J 2006; 47: 897–906. doi: 10.1111/j.1365-313X.2006.02841.x PMID: 16899084

84. Wang J, Tian L, Lee H-S, Wei NE, Jiang H, Watson B, et al. Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. Genetics 2006; 172: 507–517. doi: 10.1534/genetics.105.047894 PMID: 16172500

85. Chen ZJ, Pikaard CS. Transcriptional analysis of nucleolar dominance in polyploid plants: Biased expression/silencing of progenitor rRNA genes is developmentally regulated in *Brassica*. PNAS 1997; 94: 3442–3447. doi: 10.1073/pnas.94.7.3442 PMID: 9096413

86. Grover CE, Gallagher JP, Szadkowski EP, Yoo MJ, Flagel LE, Wendel JF. Homoeolog expression bias and expression level dominance in allopolyploids. New Phytol 2012; 196: 966–971. doi: 10.1111/j.1469-8137.2012.04365.x PMID: 23033870

87. Chaudhary B, Flagel L, Stupar RM, Udall JA, Verma N, Springer NM, et al. Reciprocal silencing, transcriptional bias and functional divergence of homeologs in polyploid cotton (*Gossypium*). Genetics 2009; 182: 503–517. doi: 10.1534/genetics.109.102608 PMID: 19363125

88. Sehrish T, Symonds VV, Soltis DE, Soltis PS, Tate JA. Gene silencing via DNA methylation in naturally occurring *Tragopogon miscellus* (Asteraceae) allopolyploids. BMC Genomics 2014; 15: 701. doi: 10.1186/1471-2164-15-701 PMID: 25145399

89. Barr CM, Fishman L. The nuclear component of a cytonuclear hybrid incompatibility in *Mimulus* maps to a cluster of pentatricopeptide repeat genes. Genetics 2011; 184: 455–465. doi: 10.1534/genetics.109.108175

90. Soltis PS, Plunkett GM, Novak SJ, Soltis DE. Genetic-variation in *Tragopogon* species—additional origins of the allotetraploids *T*. *mirus* and *T*. *miscellus* (Compositae). Am J Bot 1995; 82: 1329–1341. doi: 10.2307/2446255