POLYPLOID GENOME EVOLUTION

A thesis presented in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Plant Biology

Institute of Fundamental Sciences
Massey University
Palmerston North
New Zealand

Tina Sehrish

2014

ABSTRACT

Genome duplication is a major force influencing plant genome evolution. Many plant species have shown multiple rounds of whole genome duplications in the past. Duplicated genes show variable rate of retention, silencing, subfunctionalization and neofunctionalization which are pronounced outcomes of genome duplication. This thesis addresses polyploid genome evolution focusing on the genetic and epigenetic consequences of genome duplication. *Tragopogon dubius*, *T. pratensis* and *T. porrifolius* (diploid progenitors) and their polyploids *T. miscellus* and *T. mirus* were employed as an ideal system to examine the outcomes of polyploidy. An investigation of cytonuclear coordination in *T. miscellus* polyploids showed a maternal influence which was evident from the biased retention and expression of the maternally inherited homeolog of *rbcS* possibly to facilitate its interaction with the maternally derived *rbcL* in independently formed *T. miscellus* natural polyploids. The second study involved the genetic characterization of synthetic *T. miscellus* and *T. mirus* polyploids in the context of their relationship with each other. Results showed the presence of the same multilocus genotypes reported previously in natural *T. miscellus* and *T. mirus* and also suggested that there are certain genetic rules to the formation of polyploids; that is, only some progenitor genotypes are successful in producing polyploids. In the third study, a comparative transcriptome analysis of the reciprocally formed synthetic and natural *T. miscellus* polyploids was conducted. This study demonstrated additivity in the expression of progenitor orthologs of floral identity genes in reciprocally formed *T. miscellus* polyploids, suggesting other genetic factors are responsible for the differing inflorescence and flora morphologies in *T. miscellus*. The fourth study explored the epigenetic consequences of polyploidy. The DNA methylation status of homeologous loci previously reported to be silenced in *T. miscellus* natural polyploids was investigated. This study revealed silencing of two out of five homeologous loci by DNA methylation, suggesting

other mechanisms may be responsible for silencing of the remaining three homeologous loci. In short, collectively these studies significantly contribute to our knowledge of polyploid genome evolution in *Tragopogon* in particular and in plants in general.

ACKNOWLEDGEMENTS

This thesis marks the end of my journey of perusing a Doctoral degree. Though only my name appears on the cover of this dissertation, many people have contributed to its production and completion. I owe my deepest gratitude to Almighty Allah Who has complete knowledge about everything, for blessing me a smart brain and enabling me achieving my big goal of life i.e. PhD. I couldn't see Him but I've felt him everywhere and in every difficult moment of my life, standing to guide and help me.

I would never have been able to finish my dissertation without the excellent guidance, patience, kind help and a great support of my supervisor Dr. Jennifer A. Tate, an always welcoming, smiling and my favourite personality. I always found her there whenever I came across a research question, need technical assistance in planning an experiment and/or to discuss results or even share personal experiences of expecting and raising a baby along with research. I have been amazingly fortunate to have an advisor who gave me the freedom to explore on my own and at the same time the guidance to keep me on the right track. I feel proud in saying that I was no less than a lucky person for working with such a motivated, encouraging and a great scientist.

Dr. Vaughan Symonds, my co-adviser, has also been a great help throughout my degree and I am fan of his critical but insightful comments and constructive criticisms at different stages of my research. It was a great opportunity to work with him who made understanding of population genetics and all the involved experimentations and their results a trivial thing for me.

I take this opportunity to sincerely acknowledge the Higher Education Commission (HEC) Pakistan, for supporting my stay and studies at Massey University, New Zealand. Massey University exclusively Institute of Fundamental Sciences (IFS) deserves my sincere expression of thanks for providing highly equipped lab facilities, qualified staff and friendly & research conducive environment.

I am indebted to all my lab fellows especially Cindy Skema for taking me through high quality RNA extractions needed for MiSeq & HiSeq RNA sequencing; Prashant Joshi for technical and lab help and Rebecca Bloomer for her moral and academic support throughout my PhD candidature.

I find no words to express my gratitude to my parents for their love, care, time, patience, moral & financial support and encouragement to do more and more. Whatever I am today is the result of continuous support, dedication and prayers of my parents. I can't pay you back for the care which you have taken for my daughter "Ayesha" up to first nine months after her birth.

I am highly grateful to Muhammad Faisal, my husband and also a Ph.D scholar was the greatest motivation for my higher studies. His intellectual thoughts, love for science and encouragement always helped me to achieve my dream of life. I appreciate his patience and permission to leave him for the first year of my studies in New Zealand.

I find no one deserving for dedication this dissertation more than my cuddly, fairy and the loveliest daughter "Ayesha Bint e Sehrish Faisal" Who couldn't enjoy the warmth of her mother for first nine months for being away from me and missing my company, love, care, and proper time that she deserved.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

A	Adenosine
AG	*AGAMOUS*
AS	Antisense
AP1	*APETALA1*
bp	Base Pair
BS-Converted	Bisulfite converted
°C	Degrees celcius
CAPS	Cleaved amplified polymorphic sequence
cDNA	Complementary DNA
CTAB	Cetyl trimethylammonium bromide
Contig	Contiguous sequence
CYC	*CYCLOIDEA*
C	Cytosine
DEF	*DEFICIENS*
dNTP	Deoxynucleoside 5'-triphosphate
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
EST	Expressed sequence tags
F_1	First hybrid generation
FUE	Far upstream elements
gDNA	Genomic DNA
GLO	*GLOBOSA*
g	Gram
G	Guanine
HS	Homeolog specific
ID	Identification number
Indel	Insertion or deletion
Kb	Kilobase
♀	Maternal parent
MADS	MCM1, Agamous, Deficiens, Serum response factor (family of transcription factors with conserved DNA binding site)
Mb	Megabase
m NA	Messenger RNA
µg	Microgram
µl	Microliter
µM	Micromolar
ml	Milliliter
mM	Millimolar
MYA	Million years ago
min	Minute(s)
M	Molar
ng	Nanogram
NUE	Near upstream elements
♂	Paternal parent
PI	*PISTILLATA*

PCR	Polymerase chain reaction
PLACE	Plant cis-acting regulatory DNA elements
PVP	Polyvinyl pyrrolidone
3’ RACE	Rapid amplification of cDNA ends
RNA	Ribonucleic acid
RNase	Ribnuclease
rRNA	Ribosomal RNA
RuBisCO	Ribulose-1,5-bisphosphate carboxylase oxygenase
rbcL	Ribulose bisphosphate carboxylase large subunit
rbcS	Ribulose bisphosphate carboxylase small subunit
RIN	RNA integrity number
S_0	First generation of synthetic polyploids
S_1	Second generation of synthetic polyploids
S	sense
SEP	*SEPALLATA*
SNP	Single nucleotide polymorphism
Spp.	Species
T	Thymine
TDF	Transcript derived fragment
TE	Transposable elements
TSS	Transcription start site
U	Unit(s)
UTR	Un-translated region
U	Uracil
WGD	Whole genome duplication

CHAPTER ONE

1. Introduction

1.1 Polyploidy

Polyploidy is the heritable state of having more than two complete sets of chromosomes. The majority of polyploids possess an even number of sets of chromosomes, with tetraploidy being the most common (Jiao *et al.*, 2011; Chester *et al.*, 2012; Wendel *et al.*, 2012; Madlung, 2013). Polyploids are frequent among plants and also common among fish and amphibians (Mable *et al.*, 2011). Polyploidy in plants was first reported by Hugo de Vries on *Oenothera lamarkiana* mut. *Gigas* (*Onagraceae*), which was determined to be a tetraploid (Vaughan, 1906; Lutz, 1907). Polyploidization has been inferred to occur at least once in the evolutionary history of most species (Blanc & Wolfe, 2004a). The study of eukaryotic genomes is offering astounding evidence of the evolutionary potential of polyploids: many sequenced genomes exhibit evidence of polyploid ancestry (Soltis & Soltis, 1995; Chen *et al.*, 2007; Soltis *et al.*, 2009). In this chapter, I present a review of polyploidy, focusing on the genetic and epigenetic consequences of polyploidy observed in various plant polyploids.

1.2 Types of polyploids

Polyploids are characterized by their chromosomal composition and their mode of formation. Several different types of polyploids have been described including autopolyploids, segmental allopolyploids, true or genomic polyploids and autoallopolyploids (Stebbins, 1950; reviewed in Tate *et al.*, 2005). Primarily, two types of polyploids occur: autopolyploids and allopolyploids. Typically, autopolyploids are formed within populations of the same species (same origin), and allopolyploids result from interspecific hybridization (diverse origin) (Ramsey & Schemske, 1998; Comai, 2005a). Segmental polyploids fall between auto- and allopolyploids because they are formed within

a species, but have partial homology of their chromosomes, which exhibit both multivalent and bivalent formation; for instance $B_1B_1 \times B_2B_2 \rightarrow B_1B_1B_2B_2$. Autoallopolyploids possess both auto- and allopolyploid-derived chromosome sets, such as AAAA × BB → AAAABB (Stebbins, 1947; Swaminathan, 1954; Sybenga, 1996; Ramsey & Schemske, 1998). Aneuploids possess either an extra chromosome or are missing a chromosome resulting in a different haploid number (Myers & Hill, 1940). The term paleopolyploid refers to ancient polyploids that have been diploidized over time and the term neopolyploid refers to newly formed allo- and autopolyploids (Wolfe, 2001; Ramsey & Schemske, 2002).

1.3 Chromosomal pairing behavior of polyploids

Pairing of chromosomes during meiosis varies according to the type of polyploid. In some allopolyploids, homologous chromosomes from one genome type preferentially pair with each other at metaphase I, resulting in the formation of bivalents. While autopolyploids have more than one pair of homologous chromosomes, so pairing can take place either between two randomly (non-preferentially) selected homologous chromosomes (forming bivalents) or between more than two homologous chromosomes resulting in multivalent formation (Sybenga, 1996; Hauber *et al.*, 1999; Wu *et al.*, 2001). In polyploids, pairing affinity also depends on the degree of chromosomal homology; identical homologous chromosomes tend to pair preferentially compared to less similar chromosomes (Doyle, 1963; Benavente & Orellana, 1991). During meiosis, homologous chromosomes physically link by synapsis and recombination to ensure accurate segregation. This meiotic recombination results in either cross-overs (reciprocal exchange of DNA fragments between pairing partners) or gene conversions (unidirectional transfer of DNA fragments from one non-sister chromatid to another) (Bhalla & Dernburg, 2008). Homologous recombination (recombination between homologous chromosomes) occurs more commonly compared to homeologous recombination (recombination between distinct but related chromosomes). In allopolyploids, homeologous recombination can

lead to chromosomal rearrangements and hence genetic variability needed for adaptation (Pinto *et al.*, 2005; Nicolas *et al.*, 2007; Modliszewski & Willis, 2014).

1.4 Formation of polyploids

In general, polyploids result from mitotic or meiotic mishaps, such as failure of chromosomes to segregate, resulting in the production of either polyploid somatic tissues on a normally diploid plant or meiotic nuclear restitution during gamete formation results in unreduced gametes (2*n*) giving rise to polyploid plants. The latter is considered the more common mechanism of polyploid formation (Harlan & Dewet, 1975; Thompson & Lumaret, 1992; Brownfield & Kohler, 2011).

Somatic doubling is a rarely reported mechanism of polyploid formation, which is known to form mixoploid chimeras in meristemic tissue of sporophytes. For example, the development of a tetraploid shoot was observed on a diploid F_1 hybrid produced between *Mimulus nelsoni* and *M. lewissii* (Hiesey *et al.*, 1971). Somatic doubling is known to occur in non-meristematic plant tissues as well, where cells can initiate new growth in tumors or wounds and are an important source of new shoots (D'amato, 1952). Somatic doubling also occurs in zygotes or young embryos producing polyploid sporophytes. There is very little information available on the frequency of somatic doubling in plants and none of the effects of interspecific hybridization are known on its occurrence (reviewed in Ramsey & Schemske, 1998).

The more common mechanism of polyploid formation is by gametic non-reduction during micro- and mega-gametogenesis. The process results in the formation of 2*n* gametes, containing the complete somatic chromosome number. These 2*n* gametes are produced by meiotic events, such as cytological alterations associated with first division restitution (FDR) and second division restitution (SDR) (Ramanna, 1979; Carputo *et al.*, 2000). Formation of 2*n* pollen results from the disorientation of spindle fibres at metaphase II or abnormal cytokinesis, while 2*n* egg production is caused by the absence of

cytokinesis after telophase II, but the absence of the first and second meiotic divisions has also been reported (Werner & Peloquin, 1991; Barcaccia *et al.*, 2003). Most reports on unreduced gametes are based on pollen because it is easy to identify compared to unreduced eggs that are much harder to study (Ramsey & Schemske, 1998). The formation of polyploid embryos may occur by the union of unreduced (2*n*) gametes (Bretagnolle & Thompson, 1995; Comai, 2005a; De Storme & Geelen, 2013).

In addition to genetic factors, there are also environmental factors that can stimulate unreduced gamete formation; these include rapid changes in temperature (heat or cold treatment), x-rays, UV light, dehydration and infections. For example, *Rosa* plants were exposed to different temperature regimes. Extreme temperatures were associated with abnormal meiosis, resulting in disorientation of spindles at telophase II and an increase in the ploidy level of pollen grains (Pecrix *et al.*, 2011). Similarly, *Brassica* allopolyploids showed an increase in the frequency of unreduced gamete formation at cold temperatures (Mason *et al.*, 2011). Hence, adverse conditions could facilitate polyploid formation in the wild (Sax, 1936).

1.5 Phenotypic consequences of polyploidy

Allopolyploidy is one of the major forces involved in plant speciation, resulting from the union of two or more diverse, but generally closely associated, genomes into the same nucleus by hybridization. The accumulation of increased genetic variation in allopolyploids is possible by gene redundancy, which provides the likelihood of generating novel functional diversity between homeologous genes and genomes (Adams & Wendel, 2005b; Soltis *et al.*, 2009; Madlung, 2013).

Polyploidy can have instant phenotypic consequences, for instance, enlarged cell size and organ size, and occasionally better vigour and increased biomass (Soltis & Soltis, 1995; Comai *et al.*, 2000; Balao *et al.*, 2011). The emergence of these new phenotypes in polyploids possibly involves modifications in gene expression (Osborn *et al.*, 2003; Paun *et*

al., 2011). Polyploids often show novel phenotypes or show increased variation compared to their parents (Ramsey & Schemske, 2002). According to an evolutionary or ecological perspective, polyploid events may be observed as a stimulus for novel phenotypic changes. For example, studies revealed changes in *Brassica* for a number of environmentally critical phenotypic characters, including flowering time (Pires *et al.*, 2004b), leaf morphology, and seed set (Doyle *et al.*, 2008). Some of these characters, like drought tolerance, flowering time, pest resistance, apomixis, and increased biomass, could permit polyploids to adapt to new places and environments or increase their probability to be utilized in agriculture (Fawcett *et al.*, 2009; Van Laere *et al.*, 2011; Martin & Husband, 2012; Hannweg *et al.*, 2012).

1.6 Genetic and epigenetic consequences of polyploidy

The duplication and merger of distinct genomes in one nucleus can lead to considerable genetic and epigenetic restructuring of the duplicated genomes. The genes duplicated as a result of allopolyploidy (homeologs) have a number of evolutionary outcomes. Three potential fates are expected for these duplicated genes: (I) both copies are preserved and stay functional, (II) one copy maintains the actual function whereas the other copy is silenced or lost, or (III) subfunctionalization occurs when the two copies deviate such that each copy contributes only a part of the original gene function, or neofunctionalization may occur in which one copy attains a novel function (Ramsey & Schemske, 1998; Lynch & Conery, 2000; Edger & Pires, 2009; Roulin *et al.*, 2013).

Genetic changes post-polyploidization comprise chromosomal reshuffling, translocations, gene loss, concerted evolution of rDNA repeats, and/or transcriptomic changes (Kovarik *et al.*, 2004; Anssour *et al.*, 2009; Buggs *et al.*, 2010b; Jackson & Chen, 2010; Malinska *et al.*, 2010; Buggs *et al.*, 2012a; Chester *et al.*, 2012; Tang *et al.*, 2012; Ma *et al.*, 2013). Epigenetic changes involve DNA methylation, histone modifications, deacetylation, microRNAs and prions (Halfmann & Lindquist, 2010; Vanyushin & Ashapkin,

2011; Lee & Shin, 2012). These genetic and epigenetic modifications can occur instantly in the first generation after polyploidization or during many generations after polyploid formation (Madlung *et al.*, 2002; Adams & Wendel, 2005a; Chantret *et al.*, 2005; Skalicka *et al.*, 2005; Otto, 2007; Doyle *et al.*, 2008; Hegarty & Hiscock, 2008; Chague *et al.*, 2010; Flagel & Wendel, 2010; Dong & Adams, 2011b; Hu *et al.*, 2013; Madlung, 2013). Genetic and epigenetic outcomes of polyploidy studied in various plant genera are as follows:

1.6.1 Change in genome size

Presumably, polyploids are expected to be additive of their progenitor genomes but deviation from additivity in genome size is one of the significant consequences of a polyploidization event. Although additivity in genome size has been observed in newly formed natural and synthetic polyploids (Liu & Wendel, 2002; Pires *et al.*, 2004a; Russell *et al.*, 2013), patterns of non-additivity in polyploid genome size have been found in older and long-established polyploids (*Gossypium*, Kadir, 1976; *Vigna*, Parida *et al.*, 1990; *Nicotiana*, Leitch *et al.*, 2008). Some polyploids have been described with an increased quantity of DNA relative to their progenitors (Jakob *et al.*, 2004; Leitch *et al.*, 2008), but the overall trend in polyploid genome evolution in angiosperms is towards genome downsizing (loss of DNA following polyploidization), most probably to reduce genetic instability caused by genetic redundancy and any phenotypic effects of having larger nucleus and cell size. C values (Quantity of DNA in the gametic nucleus) in some polyploids were less than predicted (Parida *et al.*, 1990; Leitch & Bennett, 2004; Castro *et al.*, 2012; Wong & Murray, 2012; Duchoslav *et al.*, 2013; Li *et al.*, 2013). Predominantly, loss of repetitive DNA clusters like *Ty3-gypsy* and *Ty1-copia* retroelements have caused reduction in the predicted genome size in *Nicotiana* polyploids (Renny-Byfield *et al.*, 2011; Renny-Byfield *et al.*, 2013). In addition to changes in the DNA sequence, epigenetic mechanisms are also associated with transposon elements and are involved in genome restructuring and downsizing in *Spartina* and *Dactylorhiza* (Parisod *et al.*, 2009; Paun *et al.*, 2010).

1.6.2 Gene loss

Loss of one set of duplicated genes (homeologs) has been widely reported in polyploid species (Kashkush *et al.*, 2002b; Mun *et al.*, 2009; Tate *et al.*, 2009a; Buggs *et al.*, 2012a; Buggs *et al.*, 2012b; Mlinarec *et al.*, 2012; Akhunov *et al.*, 2013). Synthetic allopolyploids of *Brassica* showed extensive loss of parental DNA fragments at the F_5 generation (Song *et al.*, 1995). In a similar study of 49 independently resynthesized *Brassica* lines, homeolog losses were observed in the S_0 generation (Lukens *et al.*, 2006). Moreover, in successive generations, a number of sequence losses occurred because of homeologous recombination leading to non-reciprocal translocations (Pires *et al.*, 2004b). Recently, Xu *et al.* (2012) revealed genetic changes involving deletions and insertions of novel fragments in two independently generated sets of *Brassica napus* synthetic allopolyploids. Sequence loss (loss of anonymous DNA fragments) following polyploidization has also resulted in phenotypic variation in wheat. Elimination of parental fragments was illustrated in synthetic wheat allopolyploids and their relatives (Feldman *et al.*, 1997; Liu *et al.*, 1998), both instantaneously after formation of the polyploid (synthetic allotetraploids, Shaked *et al.*, 2001; Kashkush *et al.*, 2002b), and in generations afterwards (synthetic allohexaploids, Ma *et al.*, 2004; Ma & Gustafson, 2008). In many cases, DNA fragments appear to have been lost via homeologous recombination in the synthetic allopolyploids (Shaked *et al.*, 2001; Kashkush *et al.*, 2002a; Gaeta & Pires, 2010).

Extensive loss of the duplicated genes may be responsible for significant variation among related plant species (Paterson *et al.*, 2004; Pellicer *et al.*, 2010). One of the stimulating features of variable retention of duplicated genes involves the pattern of sequence elimination versus survivorship (Liu & Davis, 2011). Considering the functional importance of retained duplicates, it is proposed that the probability of duplicate retention was related to the number of functional interactions between the gene products. Dosage sensitive genes which are involved in various regulatory networks (like transcription factors) are retained significantly compared to other non-functional DNA sequences (Udall

& Wendel, 2006; Edger & Pires, 2009; Severin *et al.*, 2011; Wang *et al.*, 2011; De Smet & Van De Peer, 2012). For instance, in *Arabidopsis*, genes that were involved in regulatory networks such as signal transduction and transcription remain duplicated, while others involved in DNA repair were reduced to single copy (Blanc & Wolfe, 2004b; Chapman *et al.*, 2006). In contrast to the above pattern, in Asteraceae, genes associated with structural components and cellular organization were reported to be preferentially retained while genes associated with transcription and regulatory pathways were considerably underrepresented (Barker *et al.*, 2008). The chromosomal position of conserved versus lost genes, including the level to which retained genes are grouped together, is also of interest. For instance, in *Brassica napus*, the rate of sequence loss has been shown to increase with increasing genetic distance from the centromere (Nicolas *et al.*, 2012).

1.6.3 Changes in the transcriptome and proteome

Generally, it is assumed that polyploids will additively express their paternal genes but polyploidy events have significant effects on duplicate gene expression, with up- or down-regulation of one of the parental homeologs commonly resulting. Advances in high-throughput technologies have made it possible to analyze evolutionary outcomes of genome duplication events at the level of the transcriptome and proteome (De Smet & Van De Peer, 2012). Absolute comparisons between transcriptome and proteome profiles are not possible because they are not equivalent to each other due to the involvement of various post-transcriptional and post-translational modifications (Gygi *et al.*, 1999). Because the prediction of proteome expression from mRNA expression is not possible, the analysis of both transcriptome and proteome would improve our understanding of polyploid genome evolution and adaptation.

One of the most significant findings with regard to gene expression in polyploids is the unequal contribution of homeologous genes to the transcriptome, as shown in cotton (Adams *et al.*, 2003b). Adams *et al.* (2003) showed that 10 out of 40 homeologs from the A and D genomes of allotetraploid cotton demonstrate biased expression, including various

examples of reciprocal silencing among neighbouring floral whorls. This illustration of unequal contributions of duplicated genes to the transcriptome has been confirmed and extended in a number of later studies. For instance, Hovav *et al.* (2008) examined homeolog ratios for around 1400 gene pairs throughout the growth of the cotton "fiber." Biases in homeologous expression were extended to the temporal variation across developmental phases (Doyle *et al.*, 2008; Hovav *et al.*, 2008; Chaudhary *et al.*, 2009; Grover *et al.*, 2012). Hegarty *et al.* (2005) observed substantial variation in the expression levels of various genes within diploid, allohexaploid and triploid *Senecio* species. Analysis of the transcriptome of synthetic polyploids of *Arabidopsis* have shown alterations in gene expression that were associated with modified DNA methylation patterns (Yu *et al.*, 2010). Recently, analysis of leaf transcriptomes of F_1 hybrids, synthetic and natural *Gossypium* polyploids demonstrated considerable up- and down-regulation of gene expression and biases in expression towards one of the parental genomes, suggesting various regulatory and epigenetic interactions arising through transcriptome networks post-polyploidization (Dong & Adams, 2011a; Yoo *et al.*, 2013). In wheat, *Nicotiana* and *Brassica*, RNA sequencing revealed structural and functional modifications in the duplicated genes owing to neo/subfunctionalization mechanisms (Pont *et al.*, 2011; Bombarely *et al.*, 2012; Higgins *et al.*, 2012)

Koh *et al.* (2012) examined the proteome of F_1 hybrids, synthetic and natural *Tragopogon mirus* polyploids and its diploid progenitors (*T. dubius* and *T. porrifolius*) and reported expression changes after hybridization and polyploidization events and also found proteins with novel expression. Moreover, proteome analysis of early generations (F_1-F_4) of *Brassica napus* synthetic polyploids showed non-additive repatterning of their protein expression profile. However, this non-additivity did not involve expression profiles of housekeeping genes involved in regulatory networks. Hence, after polyploid formation gene silencing is a continued phenomenon and can be activated at any generation (Kong *et al.*, 2011).

1.6.4 DNA methylation

Allopolyploidy frequently causes unpredicted deviations from expected transcriptomic additivity, which may be the result of extensive epigenetic alterations. DNA methylation is one of those epigenetic modifications causing silencing of duplicated genes by methylation of the promoter or coding sequences of the genes, thereby repressing transcription by inhibiting binding of the transcription factors to the promoter (Finnegan *et al.*, 1998; Salmon *et al.*, 2005; Salmon & Ainouche, 2010). DNA methylation is the addition of a methyl group at position 5 of pyrimidine ring of a cytosine residue (Finnegan *et al.*, 1998; Chan *et al.*, 2005; Vanyushin, 2006). Cytosine methylation is important for genomic stability and is involved in genomic imprinting, transposon silencing and epigenetic regulation of gene transcription (Martienssen & Colot, 2001; He *et al.*, 2011; Vanyushin & Ashapkin, 2011; Ji & Chen, 2012). In wheat, cytosine methylation has been observed as an instant response to genome duplication (Shaked *et al.*, 2001). Similarly, Lukens *et al.* (2006) reported cytosine methylation in resynthesized *Brassica napus* allopolyploids. A high level of DNA methylation in *Brassica oleracea* has been proposed to be associated with genome plasticity and a high level of phenotypic variability (Salmon *et al.*, 2008). Madlung *et al.* (2002) observed that the phenotypic instability of synthetic *Arabidopsis* polyploids was due to an irregular methylation status causing an altered rate of transcription resulting in both gene silencing and gene activation, with the latter caused by de-methylation. Verhoeven *et al.* (2010) reported de novo methylation in triploid dandelion lineages post-polyploidization. DNA methylation is proposed to generate phenotypic variation and thereby could facilitate the adaptation of naturally occurring polyploids (Richards *et al.*, 2008). Genome wide mapping of DNA methylation in *Arabidopsis thaliana* revealed the widespread prevalence of DNA methylation in intergenic regions and repetitive sequences, while limited occurrence in the regulatory regions such as promoter, thus serving to maintain genomic integrity (Zhang *et al.*, 2006; Weber & Schuebeler, 2007; Zilberman *et al.*, 2007). Melamed-Bessudo and Levy (2012) studied the

role of DNA methylation in chromosomal recombination by employing decreased DNA methylation (*ddm1*) mutant lines in *Arabidopsis*. They reported that a decrease in DNA methylation increases the rate of meiotic crossing over in the euchromatic regions, but not in heterochromatin, suggesting a repressive role of DNA methylation in meiotic recombination. In *Oryza sativa*, the methylation level (the proportion of methylated CpG sites) of the coding regions of genes was estimated in five plant tissues for different gene categories including transposable element-like genes, duplicated genes and singleton genes (Wang *et al.*, 2013b). They reported a low level of methylation associated with high levels of expression, in contrast to a high level of methylation associated with low levels of expression suggesting genome-wide methylation is related to duplicate gene expression and determines the survivorship of duplicated genes (Wang *et al.*, 2013b).

1.6.5 Transposon activation, small RNAs and RNAi

Transposable elements (TEs) are discrete fragments of DNA that can transpose or 'jump' around the genome from one site to another unrelated site. Their random insertions in genomic sequences can mutate host DNA thereby affecting gene expression. TEs play a key role in restructuring genomes following allopolyploidization (Kidwell, 2005; Parisod *et al.*, 2010). Analysis of various allopolyploid systems has shown structural, functional and epigenetic modification by transposons during allopolyploidization (Kraitshtein *et al.*, 2010; Parisod *et al.*, 2010; Yaakov & Kashkush, 2011; Zhang *et al.*, 2013). For example, transposon activation has been observed in wheat allopolyploids. The activation of some transposable elements can lead to silencing of neighbouring downstream genes (Kashkush *et al.*, 2003; Domingues *et al.*, 2012). For instance, activation of the *Wis* transposon resulted in inactivation of the downstream *purB* gene, which is involved in softening of endosperm of the wheat kernal (Kashkush *et al.*, 2003; Chen *et al.*, 2010). After polyploid formation, epigenetic remodelling of the genome also occurs through various regulatory pathways such as RNA interfence (RNAi). RNAi helps in overcoming the gene redundancy and maintaining genomic integrity post-polyploidization (Lawrence & Pikaard, 2003). During

the process of RNAi, small interferring RNA (siRNA) molecules of 21-24 nucleotides are produced from double stranded endogenous RNAs which are formed by back-folding to form hair-pin structures. These siRNA unwind to form single-stranded siRNA molecules which integrate into RNA induced silencing complex (RISC). Then they bind to a complementary site of the target mRNA sequence and induce its cleavage by argonaute proteins, resulting in the repression of gene expression of the target mRNA (Mocellin & Provenzano, 2004). Ha *et al.* (2009) proposed that the loss of siRNA in F_1 hybrids of *Arabidopsis* was related to genomic instability, while non-additive expression of miRNA and siRNA in the *Arabidopsis* allopolyploids was associated with chromatin maintenance, vigor and adaptation. Small RNAs play significant roles in many biological processes, including the regulation of gene expression, developmental timing, the maintenance of genomic integrity and defence against invasive nucleic acids (e.g., viruses and transposons) (Cam, 2010; Ketting, 2011; Ng *et al.*, 2012; Feng & Guang, 2013).

1.7 Synthetic allopolyploid lines as a useful resource

Many polyploid species originated several million years ago such as wheat (13 mya), (Brandon, 2002), cotton (5-15 mya) (Wendel & Cronn, 2003) and *Brassica* (~20 mya) (Lagercrantz, 1998). Employing plant systems of different ages is important particularly to separate mechanisms that might describe the earliest stages of polyploid formation from those processes that are accountable for longer-term evolutionary variation (Adams & Wendel, 2004; Levy & Feldman, 2004; Otto, 2007). Moreover, crop polyploids, which are most often studied, have been subjected to artificial selection at some point during their history, which could be an argument for studying natural polyploids (Buggs, 2008).

Natural allopolyploids that were formed in the last century, like *Senecio cambrensis* (Ashton & Abbott, 1992), *Spartina anglica* (Baumel *et al.*, 2001), *Tragopogon miscellus* and *Tragopogon mirus* (Ownbey, 1950), offer a closer look into a variety of genome changes

and gene expression differences occurring upon allopolyploidization (Adams, 2007). However, synthetically made polyploids are excellent models for examining the instant outcomes of polyploidization, which cannot be shown in the corresponding natural systems and they offer insights into mechanisms that take place immediately upon formation (Adams & Wendel, 2005a). Because the progenitors of a synthetic polyploid are known with certainty, it can be specifically resolved whether widespread genome changes take place immediately after synthesis of the polyploid and if so, then the timing and mechanisms of genome changes can be determined (Song *et al.*, 1995).

Recently, a number of synthetic polyploids of various plant species have been developed by interspecific hybridizations between the diploid progenitors and then chromosome doubling of the F_1 hybrids (*Brassica*, Song *et al.*, 1995; *Arabidopsis*, Comai *et al.*, 2000; Cotton, Adams *et al.*, 2003a; Wheat, Ma & Gustafson, 2008; *Tragopogon*, Tate *et al.*, 2009b; *Miscanthus*, Chae *et al.*, 2012; Rice, Wang *et al.*, 2013a). Recent molecular studies of newly synthesized wheat (*Triticum spp*.) have shown sequence loss, gene expression changes, methylation, rearrangement of transposable elements and chromosomal instability after F_1 hybrid and allopolyploid formation (Ma & Gustafson, 2008; Yaakov & Kashkush, 2011; Yang *et al.*, 2011b). Investigation of synthetic polyploids of *Arabidopsis* have determined that genes duplicated by polyploidy (homeologs) can be silenced instantly or soon after polyploidy (Comai *et al.*, 2000). Madlung *et al.* (2005) found meiotic instability, transposon activation and chromosomal rearrangements in newly formed synthetic polyploids of *Arabidopsis*. Similarly, chromosomal rearrangements were reported by Pontes *et al.* (2004). *Arabidopsis* synthetic polyploids exhibited considerable structural genomic modifications including deletion of a major portion of the upper arm of chromosome II (Beaulieu *et al.*, 2009). Synthetic allopolyploids of *Senecio* have shown gene silencing and alterations in the transcriptome (Hegarty *et al.*, 2006a). A synthetic autotetraploid of *Aegilops tauschii* demonstrated alterations in the phenotype associated with modified DNA methylation pattern (Zeng *et al.*, 2012). Similarly, phenotypic variation

was observed in the synthetic polyploid of *Miscanthus* (Chae *et al.*, 2012). Synthetic polyploids of *Nicotiana* have also demonstrated considerable morphological, genetic and genomic changes post-polyploidization (Anssour *et al.*, 2009). Rice synthetic polyploids showed superiority in growth and seed set compared to parents and genomic in situ hybridization (GISH) showed some chromosomal translocations between parental genomes (Wang *et al.*, 2013a). On the other hand, synthetic polyploids of *Tragopogon miscellus* have not showed any evidence of gene silencing in the seven homeologs studied at the S_1 generation (Buggs *et al.*, 2009).

1.8 *Tragopogon* as a study system

Tragopogon is an ideal system for studying allopolyploidy. The genus has about 150 species found in Eurasia (Ownbey, 1950). Three diploid ($2n = 2x = 12$) species (*Tragopogon dubius*, *T. pratensis*, and *T. porrifolius*) were established in eastern Washington State, USA, and neighbouring Idaho in the early 1900s. Since their introduction, these species have recurrently produced two allopolyploid ($2n = 4x = 24$) species: *T. mirus* resulting from *T. dubius* × *T. porrifolius* and *T. miscellus* produced by *T. dubius* × *T. pratensis.* Natural *T. mirus* is produced only with *T. porrifolius* as the maternal parent, while *T. miscellus* is produced reciprocally in nature (Ownbey, 1950; Ownbey & Mccollum, 1954; Soltis *et al.*, 1995a; Soltis & Soltis, 1999).

Ownbey and McCollum (1954) applied conventional cytogenetic approaches to karyotype the six pairs of chromosomes in the diploid species of *Tragopogon*. In addition to studying morphological features, they also identified considerable chromosomal variation (e.g., terminal knobs and secondary constrictions) within different populations of diploid species to understand the multiple origins of the two newly formed allopolyploids. Considerable work has ben done to identify independent origins of these polyploid species (Soltis & Soltis, 1989; Soltis & Soltis, 1991; Soltis *et al.*, 1995a; Soltis & Soltis, 1999). Most recently Symonds *et al.* (2010) evaluated the recurrent formations and independent

origins of *T. mirus* and *T. miscellus* polyploids using progenitor-specific microsatellite markers, specifically to determine the genetic contribution of the parental species into each independently formed polyploid species. Considering only the *T. dubius* loci (common parent of both polyploids), three multilocus genotypes (containing five loci) were found in natural population, which likely represents the historical population structure of *T. dubius* at the time of polyploid formation. A lack of gene flow between populations was also reported, representing potential reproductive barriers between polyploid lineages.

Kovarik *et al.* (2005) examined rDNA loci of *Tragopogon* including internal and external transcribed spacer (ITS and ETS) in independently formed allopolyploids from various natural populations and diploid progenitor species. This study reported concerted evolution of rDNA repeats with *T. dubius* rDNA loci lost more frequently in both polyploids (Kovarik *et al.*, 2005). Pires *et al.* (2004a) compared genome sizes of independently formed *T. mirus* and *T. miscellus* from various populations and found additivity in genome size in *T. mirus* polyploids between parental genome sizes, while two populations of *T. miscellus* showed genome downsizing. Moreover, no evidence of major genomic rearrangement was observed when these *T. mirus* and *T. miscellus* polyploids were analyzed using four chromosomal markers by fluorescent in situ hybridization (FISH).

Ownbey and McCollum (1953) tried to resynthesize *T. mirus* and *T. miscellus* and generated a diploid F_1 generation, but they were not able to produce allopolyploid plants. Moreover, as natural *T. miscellus* is produced reciprocally from *T. pratensis* and *T. dubius*, Ownbey and McCollum were only able to effectively synthesize one F_1 hybrid with *T. pratensis* as the maternal parent. More recently, these allopolyploids (*Tragopogon mirus* and *T. miscellus)* were resynthesized by Tate *et al.* (Tate *et al.*, 2009b) using several individuals of the parental diploid species.

Several studies have shown loss of homeologs in *Tragopogon* allopolyploids. Early studies by Roose and Gottlieb (1976) and Soltis *et al.* (1995b) identified loss of genetic fragments using allozyme markers. Tate *et al.* (2006) identified loss of homeologous

fragments in two natural populations (Pullman and Moscow) of *T. miscellus* using cDNA–amplified fragment length polymorphism (AFLP), along with genomic and cDNA cleaved amplified polymorphic sequence (CAPS) analyses. The suggested mechanism for loss of homeologous loci was by recombination of the parental homeologs. Additivity of parental fragments for the same loci studied for the *T. miscellus* polyploids was found in synthetic F_1 hybrids between *T. pratensis* and *T. dubius*, demonstrating that these genomic changes are associated with polyploidization rather than hybridization (Tate *et al.*, 2006a). Another study of homeologous loci in ten natural populations of *T. miscellus* revealed homeolog losses in independently formed populations suggesting homeolog losses are potentially responsible for genome downsizing, but some homeologs were retained consistently (Tate *et al.*, 2009a). Similarly, frequent loss of homeologous loci in natural *T. mirus* polyploids has been observed by Koh *et al.* (2010). Buggs *et al.* (2009) analysed patterns of sequence loss in 10 sets of homeologs in five natural populations of *Tragopogon miscellus* and 44 synthetic allopolyploids from the S_1 generation of the same species. They found that none of the genes examined showed immediate loss or silencing in the re-synthesized first allopolyploid generation of *T. miscellus*. Buggs *et al.* (2009) suggested that loss and silencing of some homeologs begin in natural populations within 40 generations or less after the whole-genome duplication event. In this study, only a few loci were examined in *T. miscellus* and no changes were observed in gene expression in the S_1 generation, which appears to be in contrast to studies showing gene loss and silencing in other resynthesized allopolyploids (e.g. Adams *et al.*, 2003b; Hegarty *et al.*, 2005; Lukens *et al.*, 2006). Recently, using Sequenom MassARRAY iPLEX genotyping, loss of clustered duplicated genes post-polyploidization was reported in natural *T. miscellus* populations by Buggs *et al.* (2012a) suggesting that repeated reduction in the duplicated genomes occur after whole genome duplication events. Moreover, those missing genes were corresponding to the genes lost after ancient whole genome duplications in the family Asteraceae and lost clusters were mostly associated with dosage sensitive genes. Hence, the loss of homeologous loci is a

frequent phenomenon in *Tragopogon* as has been shown in other polyploid genera (Feldman *et al.*, 1997; Kashkush *et al.*, 2002a; Lukens *et al.*, 2006).

Chromosomal rearrangements were also commonly observed in *Tragopogon* polyploids post-polyploidization. Some multivalent formations were noticed by Ownbey (1950) in natural populations of *T. mirus* and *T. miscellus*. Lim *et al.* (2008) also observed multivalent formation at the S_0 and S_1 generation in synthetic *T. mirus* individuals. Moreover, they found loss of rDNA loci along with homeologous translocations in genomic *in situ* hybridization analyses. Later, extensive and recurrent chromosomal variation, including intergenomic translocations and aneuploidy were also observed in individuals from six populations of *T. miscellus* (Chester *et al.*, 2012) and recently karyotype restructuring was also reported in *T. miscellus* allopolyploids (Chester *et al.*, 2013). Evolution of rDNA loci has also been studied for both *Tragopogon mirus* and *Tragopogon miscellus* natural and synthetic polyploids. Uniparental losses of rDNA loci were reported with losses more frequent in natural polyploids than in synthetic polyploids. *T. dubius* rDNA loci are lost more frequently than either of the other parents, *T. pratensis* or *T. porrifolius.* Moreover, *T. dubius* rDNA loci were expressed at a higher level in spite of their reduced copy number (Matyasek *et al.*, 2007; Malinska *et al.*, 2010; Malinska *et al.*, 2011).

Polyploid formation involves considerable modification at the level of transcriptome and proteome, but as mentioned earlier, these are not directly correlated. Koh *et al.* (2012) analyzed the proteome of F_1 hybrids of *Tragopogon mirus*, as well as natural and synthetic polyploids with the progenitors, *T. dubius* and *T. porrifolius*. Out of 476 proteins identified, 68 proteins showed quantitatively differential expression in *T. mirus*. Differential expression of 32 proteins was associated with hybridization (as observed in F_1 hybrids), 22 protein changes were associated with polyploidization (observed in S_1 synthetic polyploids) and 14 changes occurred since the natural formation of *T. mirus* suggesting hybrization may have a greater influence on genome evolution than gene duplication.

1.9 Thesis chapters

The research presented in this thesis examined the genetic and epigenetic consequences of polyploidization in *Tragopogon* allopolyploids, with an emphasis on *Tragopogon miscellus*. A brief overview and hypothesis for each project follows:

1.9.1 Cytonuclear coordination in *T. miscellus* polyploids (Chapter 2)

The nuclear genome is biparentally inherited in hybrids and allopolyploids while cytoplasmic genomes (chloroplast and mitochondria) are typically maternally inherited. Cytonuclear coordination between the nuclear and cytoplasmic genomes is required for genomic stability and could be best studied in reciprocal hybrids and polyploids (Levin, 2003; Chen, 2007a). Naturally occurring and reciprocally formed *T. miscellus* allopolyploids present an ideal system to study this cytonuclear coordination. In this study, using the Riboluse-1, 5-Bisphosphate Carboxylase oxygenase (Rubisco) system, cytonuclear coordination between the maternally inherited *rbcL* (encoded by chloroplast genome) subunit and the biparentally inherited *rbcS* (encoded by nuclear genomes) subunit was examined in natural and synthetic *Tragopogon miscellus* allopolyploids. We hypothesized that the maternal homeolog of *rbcS* would be expressed in the *T. miscellus* polyploids to coordinate with the maternally inherited cytoplasmic subunit *rbcL*. For this purpose, the nuclear subunit *rbcS* was characterized in the diploid progenitors, *T. dubius* and *T. pratensis,* to determine, relative retention and expression of *rbcS* homeologs in *T. miscellus* polyploids.

1.9.2 Exploring genetic structure of *T. mirus* and *T. miscellus* synthetic polyploids (Chapter 3)

The genetic structure of natural polyploids was studied by Symonds *et al.* (2010) using progenitor-specific microsatellite loci. Symonds *et al.* (2010) reported a repeated pattern of three multilocus genotypes in both *T. mirus* and *T. miscellus* polyploids. Based on that study, here we hypothesized that patterns of polyploid formation follow genetic

'rules'. To address this hypothesis, the genetic structure of synthetic *T. mirus* and *T. miscellus* polyploids was examined and compared to natural polyploids. Here, we looked for that same pattern of multilocus genotypes and also discovered novel genotypes in the synthetic polyploids.

1.9.3 Comparative analysis of floral transcriptomes (Chapter 4)

Tragopogon miscellus has formed reciprocally exhibiting short-liguled (*T. dubius*♂ × *T. pratensis*♀) and long-liguled (*T. dubius*♀ × *T. pratensis*♂) forms (Soltis & Soltis, 1989). As these morphologies were repeated in synthetic polyploids, here we hypothesize a maternal influence to the formation and variation in inflorescence/floral morphology. To determine genetic factors that might be controlling these floral morphologies, floral transcriptomes were studied from diploid parents and reciprocally formed natural and synthetic *T. miscellus* polyploids. RNA sequencing data (transcriptome) were analyzed to compare variation in the transcriptome between natural and synthetic *T. miscellus* polyploids to evaluate transcript abundance and up- or down-regulation of progenitor-specific floral development genes post-polyploidy.

1.9.4 DNA methylation: A gene silencing mechanism post-polyploidization (Chapter 5)

Gene silencing of parental homeologs have been previously reported in *Tragopogon miscellus* polyploids (Tate *et al.*, 2006a; Buggs *et al.*, 2009). DNA methylation is one of the epigenetic mechanisms that results in gene silencing. We hypothesized that DNA methylation would be responsible for the silencing of previously reported homeologous loci in *Tragopogon miscellus* polyploids. To investigate this, the pattern of CpG methylation was analyzed for those previously studied five homeologous loci [S2, S3, S8, S18 and TDF-44 (Tate *et al.*, 2006a; Buggs *et al.*, 2009)] in several individuals of natural *T. miscellus* polyploids using comparative bisulfite sequencing.

CHAPTER TWO

This chapter is in preparation as:

***Sehrish T**, Symonds VV, Soltis DE, Soltis PS, and Tate JA. Biased paternal genomic loss and maternal expression of rbcS homeologs in Tragopogon miscellus (Asteraceae) allopolyploids: insight into cytonuclear compatibility using Rubisco. In prep for: New Phytologist.*

2. Biased paternal genomic loss and maternal expression of *rbcS-1* homeologs in *Tragopogon miscellus* (Asteraceae) allopolyploids: insight into cytonuclear compatibility

2.1 Abstract

- Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), which consists of a chloroplast-encoded large subunit (rbcL) and a nuclear-encoded small subunit (rbcS), was used to examine cytonuclear interactions in naturally occurring and synthetic *Tragopogon miscellus* (Asteraceae) allotetraploids (including reciprocally formed individuals). *Tragopogon miscellus* formed recently (~80 years ago) and repeatedly via hybridization and whole-genome doubling.
- Genome-specific single nucleotide polymorphisms (SNPs) in the diploid progenitors (*T. dubius* and *T. pratensis*) were used to identify patterns of inheritance, retention, and expression of *rbcL* and the duplicated *rbcS-1* gene copies (homeologs) in 25 synthetic and 78 natural *T. miscellus* individuals.
- All allopolyploids inherited the maternal *rbcL* copy. Both parental homeologs of *rbcS-1* were retained and expressed in all synthetic and most natural individuals of *T. miscellus*. Genomic loss of one homeolog was apparent in ten natural polyploids, with retention of the maternal *rbcS-1* copy more frequent. Eight natural individuals retained both homeologs, but showed biased expression, always in favor of the maternal *rbcS-1* copy.
- Cytonuclear coordination occurs relatively early following polyploidization, but it does not appear immediately upon formation in *Tragopogon miscellus*. Moreover, the pattern of biased maternal retention and expression of *rbcS-1* homeologs is repeated across independently formed populations and in reciprocally formed individuals, indicating repeatability of enhanced cytonuclear interactions.

2.2 Introduction

Allopolyploidy is a major force in plant speciation and results from the union of two or more diverse, but generally closely related, genomes by hybridization (Soltis & Soltis, 2009; Te Beest *et al.*, 2012). Genomic data indicate that all angiosperms may be regarded as polyploid, if paleopolyploid events are taken into account (Soltis *et al.*, 2009; Yang *et al.*, 2011a). Allopolyploid genomes experience both instant (immediately after the duplication event) and long-term evolutionary changes, which may involve a variety of genetic and epigenetic interactions leading to genome alteration, regulatory incompatibilities, chromosomal abnormalities, and reproductive failures (Comai *et al.*, 2000; Hegarty *et al.*, 2006b; Chen, 2007b; Gaeta *et al.*, 2007; Ma & Gustafson, 2008; Xiong *et al.*, 2011; Chester *et al.*, 2012; Feldman *et al.*, 2012; Yoo *et al.*, 2013). Polyploidy has been considered a driver of modifications in gene function, potentially resulting in any of three fates of the duplicated genes (homeologs): (I) both copies are preserved and stay functional, (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 one copy attains a novel function (neofunctionalization) (Lynch & Conery, 2000; Prince & Pickett, 2002; Blanc & Wolfe, 2004b; Conant & Wolfe, 2008; Doyle *et al.*, 2008; Edger & Pires, 2009; Roulin *et al.*, 2013).

In newly formed hybrids and allopolyploids, coordination between the maternally inherited cytoplasmic (chloroplast and mitochondrial) and the biparentally inherited nuclear genomes is required to facilitate genomic stability (Chen, 2007b). Most protein complexes consisting of multiple subunits have proteins encoded by both nuclear and organellar genomes (Rodermel *et al.*, 1988). For example, the enzyme Riboluse-1, 5-bisphosphate carboxylase/oxygenase (Rubisco) is involved in carbon fixation in the Calvin cycle (Staswick, 1994): structurally, it is multi-meric, commonly composed of eight large subunits encoded by the chloroplast gene *rbcL* and eight small subunits encoded by the nuclear *rbcS* (Miziorko & Lorimer, 1983; Clegg, 1993). Depending on the species, *rbcS* is

part of a multigene family typically ranging from 4-22 members (Krebbers *et al.*, 1988; Dean *et al.*, 1989; Spreitzer, 2003). As a result of the interaction between co-inherited organellar genomes and the nuclear genome, co-adaptation is likely to occur. For the co-adaptation of nuclear and cytoplasmic genomes in a hybrid or allopolyploid, the evolution of the nuclear gene (*rbcS*) in the hybrid/polyploid may be essentially dependent on coordination with its counterpart in the chloroplast genome (*rbcL*).

The effects of cytonuclear interactions between the nuclear genome of one species and cytoplasmic genome of the other species have been assessed by the synthesis of reciprocal hybrids (Edmands & Burton, 1999; Levin, 2003). Cytonuclear interactions are considered responsible for post-zygotic hybrid incompatibilities and speciation (Levin, 2003; Fishman & Willis, 2006) and have also caused striking differences in the floral traits in reciprocal diploid hybrids (Grant, 1956; Oehlkers, 1964). Similarly, differences in floral morphology have also been observed in reciprocally formed *Tragopogon* polyploids, which differ in ligule length (Ownbey, 1950; Ownbey & Mccollum, 1953). Cytonuclear coordination may also be a possible contributor to the directional genomic changes and preferential expression of some genes in reciprocally formed allopolyploids of *Brassica* and *Tragopogon* (Song *et al.*, 1995; Tate *et al.*, 2006c; Wolf, 2009). To date, only one study has investigated cytonuclear interactions in allopolyploids using Rubisco. In *Gossypium*, the *rbcS* gene family consisted of four copies (two long and two short) (Gong *et al.*, 2012). All of the progenitor *rbcS* genes were retained and expressed in all five *Gossypium* allopolyploid species, but leaf transcriptome analysis revealed preferential expression of maternal *rbcS* homeologs (A-homeolog) (Gong *et al.*, 2012). Furthermore, non-reciprocal homeologous recombination was detected among parental 'A' and 'D' genome copies in the polyploids, suggesting gene conversion following genome doubling. As the *Gossypium* polyploids are 1-2 Mya (Wendel & Cronn, 2003), additional studies on young polyploids are needed to understand how early cytonuclear coordination might occur.

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 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, with short-liguled (*T. dubius* ♂ × *T. pratensis* ♀) and long-liguled (*T. dubius* ♀ × *T. pratensis* ♂) forms resulting (Ownbey, 1950; Ownbey & Mccollum, 1953; Soltis & Soltis, 1999). Recurrent formation of both allopolyploids, represented by populations from different towns in the Palouse (Soltis & Soltis, 1991; Soltis & Soltis, 1995; Soltis *et al.*, 2004; Symonds *et al.*, 2010), offers an opportunity to determine if independently formed polyploids respond similarly to cytonuclear coordination. Previous studies have identified a myriad of genomic and transcriptomic modifications in the *Tragopogon* allopolyploids in the short time since their formation (less than 80 years), including differential expression of homeologous loci, homeolog loss and silencing, differential proteome expression (Tate *et al.*, 2006c; Buggs *et al.*, 2009; Tate *et al.*, 2009a; Koh *et al.*, 2010; Buggs *et al.*, 2011; Koh *et al.*, 2012), and extensive chromosomal variation, such as aneuploidy and intergenomic translocations (Lim *et al.*, 2008; Chester *et al.*, 2012). Moreover, the formation of synthetic polyploids of *Tragopogon* has allowed the analysis of genomic modifications at early stages of polyploid formation (Tate *et al.*, 2009b).

Here, we use the Rubisco system (*rbcS* and *rbcL*) to examine cytonuclear coordination in natural and synthetic *Tragopogon miscellus* allopolyploids, representing independent and reciprocal formations. We characterize *rbcS-1* in the *Tragopogon* diploid parental species to answer the following questions: (1) How divergent are the *rbcS-1* and *rbcL* homeologs of the *T. miscellus* progenitors? (2) Is there differential retention of *rbcS-1*

homeologs in *T. miscellus*? (3) Do the natural and synthetic polyploids of *T. miscellus* show equal or biased expression of the *rbcS-1* homeologs?

2.3 Materials and Methods

2.3.1 Plant material

The natural populations sampled for *Tragopogon dubius* Scopoli, *T. pratensis* L. and *T. miscellus* Ownbey are listed in Table S2.1, as are the synthetic lineages of the allotetraploid *T. miscellus* examined. To capture potential variability in the diploid progenitors genomic DNA and cDNA were included in the study for 20 *T. dubius* and *T. pratensis* individuals from six natural populations (Table S2.1). For *T. miscellus*, four synthetic lineages (25 individuals) and ten natural populations (78 individuals total: 8-12 individuals per population) were sampled. Plant material for most of the natural polyploids and diploids was the same as that used in Tate *et al.* (2006c; 2009a) and Buggs *et al.* (2009). For synthetics, mature seeds were grown under standard glasshouse conditions at Massey University (Palmerston North, New Zealand). For some *T. miscellus* populations, genomic DNA and cDNA were not available for the same individuals (Table S2.1). Plant material for the synthetic *T. miscellus* was the same as that produced by Tate *et al.* (2009b).

2.3.2 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 (Doyle & Doyle, 1987). 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).

Table 2. 1 *rbcS-1* primers designed in this study.

Primer Name	Experiments	Primer and/ 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	GTAATTCGCATCACTATAGCTCACCGCTGGTCGACGGCC CGGGCTGGT
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	CTTTGGATGGACCTCAATCG
rbcS-F1	Sequencing	CAAAACATACCCATAACGTATCAGCC
rbcS-R3	Sequencing	AGCAGAAACATAAATTTTTATTATTATCATC
AS-Pra-Snp3	Homeolog specific RT-PCR	AAGGCCAATGACTTTTCCTCCCGC
AS-Dub-Snp3A	Homeolog specific RT-PCR	AAGGCCAATGACTTTTCCTCCCAT
AS-F-Pra	Homeolog specific RT-PCR	CATATATACAGGTGTGGCCACCAAGC
AS-F-dub2	Homeolog specific RT-PCR	CATATATACAGGTGTGGCCACCAAGT

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

Full length *rbcL* was amplified from the diploid progenitors and *T. miscellus* using Asteraceae-specific primers *rbcL1* and *rbcL2* (Olmstead *et al.*, 1992; Hillis *et al.*, 1996). 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 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 (Panero & Crozier, 2003).

Initial amplification of *rbcS* was accomplished by designing PCR primers from *T. dubius* ESTs with a reference to *Lactuca sativa rbcS* sequence (AF162210) using Primer3 (Rozen & Skaletsky, 2002). Primers for one copy, hereafter *rbcS-1*, (*rbcS*-2F and *rbcS*-2R) were used for the amplification of both genomic and cDNA of the diploids (Table 2.1). A second *rbcS*-like sequence was identified, 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 full-length *rbcS-1* (Figure S2.2). For this second copy, *rbcS-2*, 5' genome walking (using methods described later for *rbcS-1*) revealed the presence of a long ~700 bp long intron-like sequence (data not shown), not found in any angiosperm group to date (Dean *et al.*, 1989). Because *rbcS-2* did not have the conserved regions typical of other rbcS subunits in angiosperms (Dean *et al*., 1989) and as found in *rbcS-1*, we did not consider this to be a functional copy and so did not examine it further. 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). 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.4.10.1. 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

natural *T. miscellus* polyploids; these longer pieces of *rbcS-1* were then sequenced using the aforementioned sequencing protocol with both forward (*rbcS*-2F) and reverse (*rbcS*-8R) primers (Table 2.1). Chromatograms were inspected to determine retention and relative expression in polyploids (Adams *et al.*, 2004). Additivity of peaks was interpreted as retention of both homeologs and/or additive expression of both homeologs. Peaks deviating <50% from additivity were interpreted as showing biased relative expression.

2.3.4 Genomic and cDNA CAPS analysis

Sequences of *rbcS-1* and *rbcL* for the diploid parents were separately aligned to determine sequence variation that could differentiate parental homeologs in *T. miscellus*. The programs dCAPS Finder 2.0 (Neff *et al.*, 2002) and NEB Cutter v.1.0 (Vincze *et al.*, 2003) were used to identify diagnostic restriction sites between parental *rbcS-1* sequences. Genomic and cDNA cleaved amplified polymorphic sequence (CAPS) analysis was performed for *rbcS-1* using forward primer *rbcS*-3F and reverse primer *rbcS*-8R designed from cloned sequences of the diploid parents (Table 2.1). The amplified region included exon 1 (from aligned position 433 bp), intron 1, and exon 2 (to position 1054 bp) (Fig. 2.1). 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 fragment sizes 272 bp, 167 bp, 154 bp and 29 bp), while *T. pratensis* is cut at two positions (resulting fragment sizes 327 bp, 272 bp and 29 bp). For 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 natural and synthetic polyploids of *T. miscellus* and digested following the same protocols. For cDNA CAPS, TDF85 was amplified and digested (Tate *et al.*, 2006) to verify equal expression of parental homeologs.

2.3.5 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) (Siebert *et al.*, 1995). Two outward 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 2.1). 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 2.1). *Tragopogon dubius* genomic DNA was digested with three different blunt-cutting enzymes: *EcoRV*, *ScaI* and *DraI* (New England Biolabs) independently in separate reaction tubes containing 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. Adapter ligation to the cleaned, digested genomic DNA was performed 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 of adapter primer AP1 (Forward) and gene-specific primer GS1 (Reverse) (Table 2.1) 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 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 for the nested 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 2.1), 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 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 2.1). First-strand cDNA from *T. dubius* was made using an oligo(dt) sequence that incorporates a 3' RACE specific primer sequence at the 5' end (3'RACE oligodT Adapter, Table 2.1). 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 of the nested primers (Inv. Fwd 1 and 3' RACE adapter primer, Table 2.1) 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 for *T. dubius*. Once we obtained the complete *rbcS-1* gene sequence for *T. dubius*, new primers were designed (*rbcS*-F1 and *rbcS*-R3, Table 2.1) for the amplification and sequencing of the complete *rbcS-1* gene from genomic DNA and cDNA of *T. pratensis.*

2.3.6 Prediction of *rbcS-1* gene structure

Gene structure of *rbcS-1* was predicted using Augustus (Version 2.6) (Stanke *et al.*, 2008) and GENSCAN (Burge & Karlin, 1997). These programs were used to confirm the transcription start site (TSS), exons, introns and other regulatory sequences that were already determined by cDNA sequencing of the complete *rbcS-1* gene. Plant Promoter Analysis Navigator (PlantPAN) (Chang *et al.*, 2008) was used to identify promoter sequences of *rbcS-1*, putative transcription factor binding sites in the promoter region, as well as conserved motifs in the promoter.

2.3.7 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 (Li *et al.*, 2004). Homeolog specificity was assured by adding a mismatch one bp away from the 3' end of each of the two forward HS primers (*T. dubius*: AS-dub-Snp3A, *T. pratensis*: AS-Pra-Snp3, Table 2.1). A common reverse primer was designed downstream of the polymorphic site (AS-R3, Table 2.1). For the present experiment, primers were designed at the third SNP in exon 1 (corresponding to position 471 bp, Fig. 2.1). 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 and 0.5 U *Taq* polymerase (New England Biolabs, USA) and 15ng/µl template (cDNA). The amount of template cDNA included in the PCR was quantified using a Nanodrop ND-1000 spectrophotometer (Thermo Fisher Scientific, USA). PCR products were run on a 1.5% agarose gel.

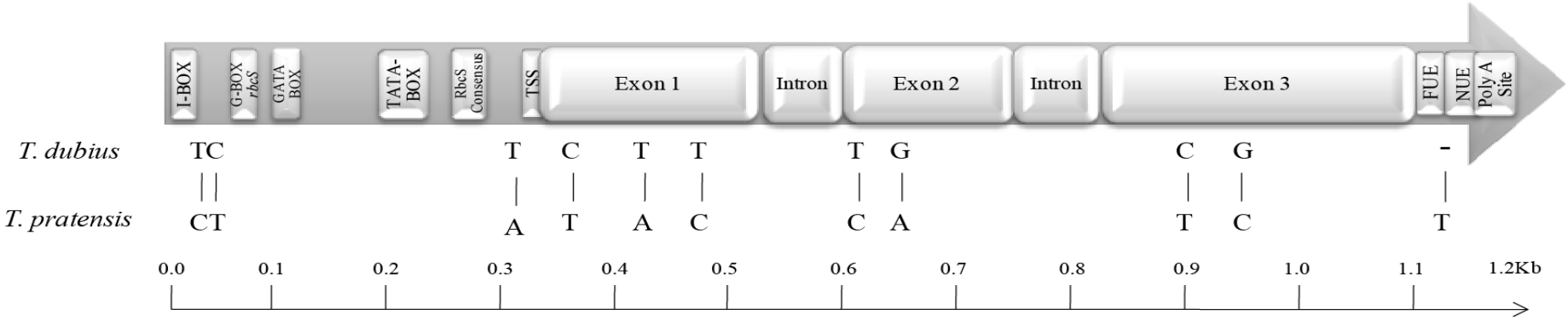


Fig. 2.1 *rbcS-1* gene structure and locations of SNPs. The structure of the *rbcS-1* gene is shown at the top with both coding regions and non-coding regulatory elements indicated. Locations for SNPs between *T. dubius* and *T. pratensis* homeologs have been scaled along the length of the *rbcS-1* gene.

2.4 Results

2.4.1 *rbcS* gene family

Two *rbcS* gene copies, *rbcS-1* and *rbcS-2* were discovered from the *T. dubius* EST database. These two *rbcS* genes are fairly divergent from each other with several SNPs, insertions and deletions in the genic regions and were even more variable at 3' UTRs. Of these two *rbcS* genes, *rbcS-1* was expressed and determined to be a completely functional copy, while the second *rbcS* gene, *rbcS-2* was considered a pseudogene and a truncated copy as it had pre-mature stop codons compared to *rbcS-1* protein sequence and 5'genome walking experiments also yielded non-*rbcS* genomic sequences at its flanking ends (Fig. S2.2). Hence, we focused on the *rbcS-1* gene to follow cytonuclear coordination in the *Tragopgon* diploids and polyploids.

2.4.2 Characterization of *rbcS-1* in *Tragopogon*

The total length of *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 *Tragopogon pratensis* (the length of *rbcS-1* sequence varied due to insertion/deletion events). For convenience, we hereafter refer to the aligned length of *rbcS-1* sequences from both species (1219bp). The 5' upstream promoter elements and 3' downstream polyadenylation signals were predicted with reference to *Arabidopsis*, tomato and tobacco using the programs PlantPAN (Chang *et al.*, 2008) and PLACE (Higo *et al.* 1999). Figure 2.1 shows the locations of the transcription start site (TSS), exons, introns and other non-coding regulatory elements. From the TSS at 331 bp, there were 3 exons of 174 bp, 133 bp and 232 bp, respectively, which were separated by two introns. The length of the first intron differed in *T. dubius* (80 bp) and *T. pratensis* (86 bp) while the second intron was 80 bp long in both progenitors. Common promoter features, such as TATA-BOX, G-Box rbcS, GATA-BOX, I-BOX and rbcS-Consensus are highlighted in Figure 2.1. The TATA-BOX was found in the *Tragopogon rbcS-1* promoter region with reference to the TATA-Box sequence of pea (*Pisum sativum*), which was

responsible for tissue-specific promoter activity in pea (Shirsat *et al.*, 1989). G-Box *rbcS* "GCCACGTGT" and I-Box are conserved sequences found in the promoter region of light-regulated genes and were located at 69 bp and 100 bp, respectively, in *Tragopogon*. These motifs are also found in the promoter of the *rbcS* gene sequence of tomato and *Arabidopsis* (Giuliano *et al.*, 1988). GATA is a conserved motif that is present in the promoter of *Petunia* chlorophyll a/b binding proteins (Lam & Chua, 1989). Four GATA repeats were found in the *Tragopogon rbcS-1* gene. The general *rbcS*-CONSENSUS "AATCCAA" was located at 272 bp in the promoter; this conserved region is present in the promoter of the *rbcS* gene in *Petunia*, tomato, tobacco and *Arabidopsis* (Manzara & Gruissem, 1988). Many other transcription factor binding sites in the *Tragopogon rbcS-1* promoter regions were also identified by PlantPAN with reference to *rbcS* gene sequence of other closely related species (Table S2.2). Downstream polyadenylation signals were also predicted. Generally, there are three conserved Poly(A) signals near mRNA ends, such as the Far-upstream Elements (FUEs), Near-upstream Elements and Cleavage site (Hunt, 1994). Poly(A) signals for the *Tragopogon rbcS-1* gene were "TTGTAA" (FUEs), "AATAAA" (NUEs) and cleavage site "TA" located at 1083 bp, 1188 bp and 1209 bp, respectively (Fig. 2.1). Poly(A) signals for the *Tragopogon rbcS-1* gene were predicted with reference to previous reports on Poly(A) signals for the *rbcS-1* gene (Hunt & Macdonald, 1989; Li & Hunt, 1997).

Genomic DNA and cDNA sequences of *rbcS-1* analyzed from all *Tragopogon dubius* individuals and *T. pratensis* individuals from various populations have not shown any intraspecific variation in their *rbcS-1* sequences.

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

Sequence analysis of *rbcS-1* exons revealed seven single nucleotide polymorphisms (SNPs) and a 1-bp indel that distinguishes *T. dubius* and *T. pratensis* homeologs (Fig. 2.1). The indel was located between polyadenylation signals at 1114 bp. Three SNPs occurred in exon 1 (at 357 bp, 424 bp, 471 bp) and two each in exon 2 (605

bp, 632 bp) and exon 3 (898 bp, 949 bp). Six of these SNPs were synonymous substitutions, with the second SNP at 424 bp a non-synonymous change between the *rbcS-1* homeologs. The non-synonymous substitution resulted 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: three SNPs in the promoter region at 31 bp, 32 bp and 322 bp, respectively, 10 SNPs and four indels (ranging in size from 1-3 bp) in the first intron and seven SNPs in the second intron. Analysis of predicted protein structure of the *rbcS-1* homeologs using the protein homology/analogy recognition engine Phyre[2] V 2.0 (Kelley & Sternberg, 2009) revealed that the non-synonymous SNP resides in an alpha-helix and does not cause any difference in 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.

To determine the pattern of retention of these subunits, genomic sequences of both *rbcS-1* and *rbcL* were analyzed from 25 synthetic polyploid individuals (representing five independently generated lineages) and 78 natural polyploids from 10 populations of *T. miscellus*. In the case of *rbcL*, all the synthetic and natural polyploids had the maternally-derived sequence (i.e., *T. pratensis* for the short-liguled form and *T. dubius* for the long-liguled form). In contrast, for *rbcS-1*, all the synthetic polyploids and most of the natural polyploids had both *T. dubius* and *T. pratensis rbcS-1* homeologs, as determined by additivity in the sequence chromatograms and genomic CAPS analysis (Fig. 2.2a). Based on genomic CAPS analysis and sequencing results, Figure 2.3a graphically shows pattern of retention of parental homeologs of *rbcS-1* in long and short-liguled natural *T. miscellus* populations. Ten natural polyploid individuals from 5 populations (Spangle, Garfield, Albion, Moscow, and Pullman) had only one homeolog present in 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*, Table 2.2, Table S2.1 and Fig. 2.3a).

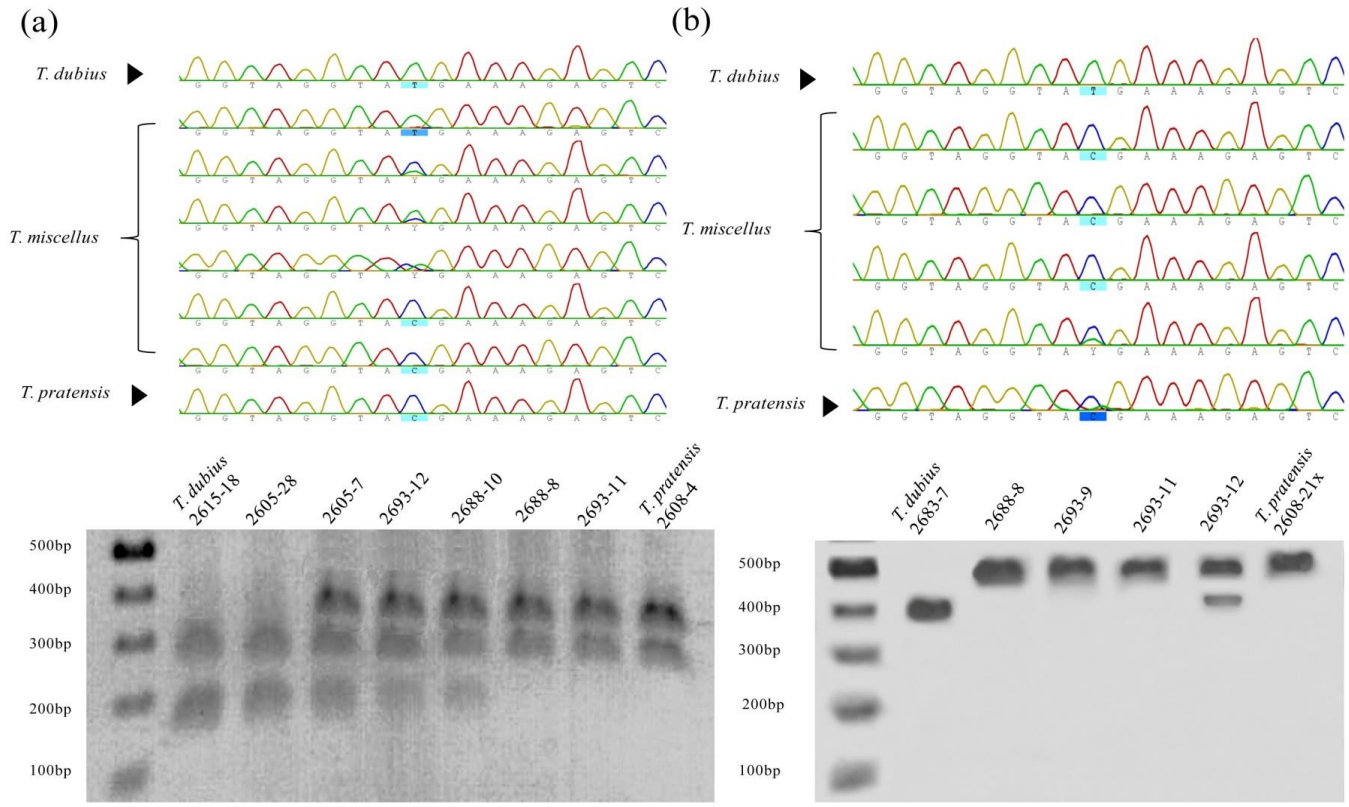


Fig. 2.2 CAPS analysis showing additivity and expression of *rbcS-1* homeologs. (a) Genomic DNA CAPS results for six representative samples and (b) cDNA CAPS results for four representative samples of naturally occurring *T. miscellus* polyploids along with representative diploid parents, *T. dubius* and *T. pratensis*. Chromatograms belong to the same polyploid samples (from top to bottom) as in the gel photo below (from left to right). Chromatograms show sequence polymorphisms at the third SNP in exon 1 for both genomic DNA and cDNA sequences of the natural polyploids. "L" following the polyploid name denotes the long-liguled form, and "S" denotes the short-liguled form.

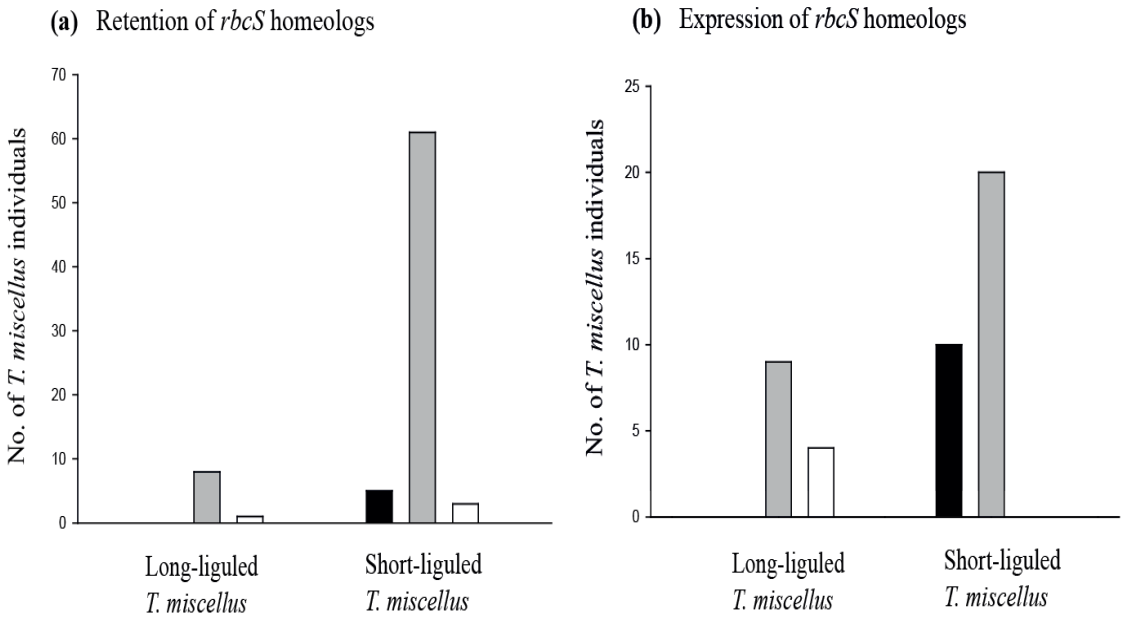


Fig. 2.3 Graphical illustration of retention and expression of parental *rbcS-1* homeologs in *T. miscellus* (short and long liguled) natural polyploids. Bar charts show (a) retention and (b) expression of parental *rbcS-1* homeologs in *T. miscellus* polyploids. Black, grey and white colours of the bars correspond to retention/expression of *T. pratensis rbcS-1* homeolog, both progenitor *rbcS-1* homeologs and *T. dubius rbcS-1* homeolog respectively. Short liguled and long liguled individuals are represented with separate bars.

One long-liguled individual from Pullman had only the *T. dubius* (maternal) genomic homeolog (Table 2.2, Fig. 2.3a).

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

Relative expression of parental *rbcS-1* homeologs was determined by examining sequence chromatograms for additivity at SNPs (Adams *et al.*, 2004) and cDNA CAPS analysis (Fig. 2.2b). For the six synthetic polyploid individuals, all showed equal expression of the parental homeologs (Table S2.1). Based on cDNA CAPS analysis and cDNA sequencing results, Figure 3b graphically shows pattern of expression of parental homeologs of *rbcS-1* in long and short-liguled natural *T. miscellus* populations. Of the 43 natural polyploid individuals examined, 14 showed deviation from additive expression

(Table 2.2, Figure 2.3b). In all cases, the maternally derived *rbcS-1* homeolog was either expressed at a greater level or was the only copy expressed (Fig. 2.2b).

Table 2.2 Naturally occurring individuals of *Tragopogon miscellus* that showed bias in the retention and expression of parental *rbcS-1* homeologs. A dash (-) indicates that we were not able to study a particular individual for both retention (genomic DNA) and expression (cDNA).

Population	Short or long liguled	Lineage	Retention of *rbcS* homeologs	Expression of *rbcS* homeologs
Spangle	Short	2693-7	Both homeologs	*T. pratensis* > *T. dubius*
Spangle	Short	2693-9	Only *T. pratensis* copy	Only *T. pratensis* copy
Spangle	Short	2693-11	Only *T. pratensis* copy	Only *T. pratensis* copy
Oakesdale	Short	2671-2	Both homeologs	*T. pratensis* > *T. dubius*
Oakesdale	Short	2671-11	Both homeologs	*T. pratensis* > *T. dubius*
Garfield	Short	2688-8	Only *T. pratensis* copy	Only *T. pratensis* copy
Garfield	Short	2688-12	Both homeologs	*T. pratensis* > *T. dubius*
Moscow	Short	2604-17	Only *T. pratensis* copy	Only *T. pratensis* copy
Moscow	Short	2604-22	Only *T. pratensis* copy	Only *T. pratensis* copy
Moscow	Short	2604-43	Both homeologs	*T. pratensis* > *T. dubius*
Albion	Short	2625-3	Only *T. dubius* copy present	-
Albion	Short	2625-6	Only *T. dubius* copy present	-
Albion	Short	2625-8	Only *T. dubius* copy present	-
Troy	Short	2682-5	Only *T. pratensis* copy present	-
Pullman	Long	2605-9	-	*T. dubius* > *T. pratensis*
Pullman	Long	2605-13	Both homeologs	*T. dubius* > *T. pratensis*
Pullman	Long	2605-28	Only *T. dubius* copy	Only *T. dubius* copy
Pullman	Long	2605-46	Both homeologs	*T. dubius* > *T. pratensis*

Ten short-liguled individuals of *T. miscellus* (i.e., with *T. pratensis* as the maternal parent) from four populations (Spangle=3, Oakesdale=2, Garfield=2, and Moscow=3) showed biased (greater) expression of the *T. pratensis* copy when both genomic homeologs were present or only expressed the *T. pratensis* copy (because the genomic copy of *T. dubius* was missing). Four long-liguled individuals (*T. dubius* as the maternal parent) showed biased (greater) expression of the maternal *T. dubius* copy (Table 2.2, Table S2.1 and Fig. 2.3b).

To verify the biased expression of the maternally derived *rbcS-1* homeolog in the natural polyploids, homeolog-specific RT-PCR was conducted. For ten polyploids, unequal expression of *rbcS-1* homeologs was estimated from the band intensity of each copy in the natural polyploid samples (Fig. 2.4). For example, three *T. miscellus* polyploids from Spangle showed expression bias in favor of the *T. pratensis* homeolog.

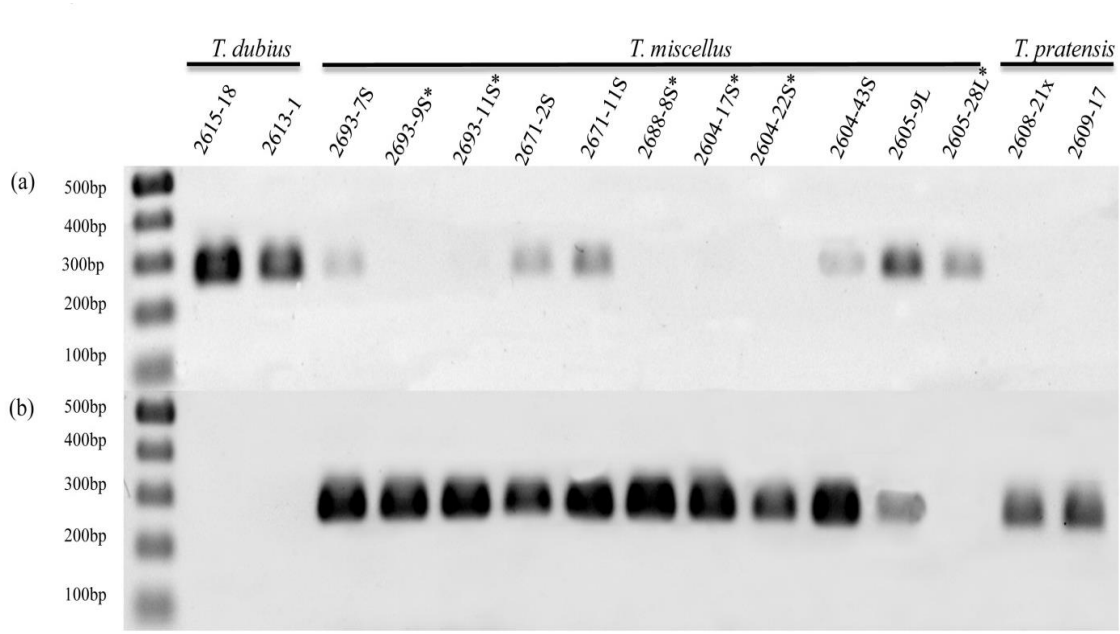


Fig. 2. 4 Homeolog-specific RT-PCR of *rbcS-1*. (a) *T. dubius* and (b) *T. pratensis* homeolog-specific RT-PCR results are shown for 11 *T. miscellus* natural polyploids with representative diploid parents, *T. dubius* and *T. pratensis*. Six individuals of the allotetraploid *T. miscellus* (indicated by an asterisk (*) showed genomic loss of one parental fragment (*T. dubius* homeolog was lost in five short-liguled polyploids, and the *T. pratensis* homeolog was lost in one long-liguled Pullman individual 2605-28). The remaining five natural polyploids show expression biased towards one of the parents. "L" following the polyploid name denotes the long-liguled form, and "S" denotes the short-liguled form.

In two of these individuals (2693-9 and 2693-11) the *T. dubius* genomic copy was lost, while the third individual (2693-7) retained both genomic copies, the *T. dubius* homeolog showed reduced expression compared to the *T. pratensis* homeolog (Fig. 2.4).

Two individuals from Oakesdale (2671-2 and 2671-11) retained both genomic homeologs, but showed reduced *T. dubius* relative expression compared to the *T. pratensis* homeolog (Fig. 2.4). In the case of the long-liguled Pullman population, one polyploid individual (2605-9) showed reduced expression of the *T. pratensis* homeolog compared to

T. dubius, while another individual (2605-28) showed genomic loss of the *T. pratensis* homeolog and therefore only expressed the *T. dubius* copy (Fig. 2.4). TDF-85 was used as a control for additive expression in the polyploids because it has been shown to be retained as duplicate in all *T. miscellus* individuals examined to date (Tate *et al.*, 2006c; Tate *et al.*, 2009a; Malinska *et al.*, 2011). RT-PCR and CAPS analysis for TDF-85 showed equal expression of both parents in the polyploid *T. miscellus* that showed biased expression for *rbcS-1* homeologs here (Figure S2.1).

2.5 Discussion

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

In higher plants, small subunit rbcS is fairly divergent and present as a multigene nuclear family compared to plastid *rbcL* which is highly conserved and present in single copy (Spreitzer & Salvucci, 2002). To explore cytonuclear coordination in young allopolyploids, here we characterize *rbcS-1* in the diploid progenitors of the allotetraploid (*T. miscellus*) by retrieving complete *rbcS-1* sequence and predicting gene and protein structure. As is the case in most other eudicots (Dean *et al.*, 1989), the *rbcS-1* gene in *Tragopogon* consists of three exons separated by two short introns. Two copies of the *rbcS* gene (*rbcS-1* and *rbcS-2*) were found in *Tragopogon,* but the second copy was not functional (see results), while in other angiosperms, the *rbcS* gene family ranges in size from four *rbcS* genes (*Arabidopsis*) to more than 22 (wheat) (Krebbers *et al.*, 1988; Sasanuma, 2001; Spreitzer, 2003). 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 (Outchkourov *et al.*, 2003; Izumi *et al.*, 2012). In the Asteraceae, *Lactuca sativa* (tribe Cichorieae) have six *rbcS* genes (Goumenaki *et al.*, 2010) and *Flaveria* (Heliantheae) contain 5-16 *rbcS* genes in any *Flaveria* species (Kapralov *et al.*, 2011), while *Helianthus* (Heliantheae)(Waksman *et al.*, 1987) and *Chrysanthemum* (Anthemideae)(Outchkourov *et al.*, 2003) each have 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 (Wolfe, 2001; Wang *et al.*, 2005).

Interspecific *rbcS-1* sequence variation was low between the parental diploids *Tragopogon dubius* and *T. pratensis* with few polymorphisms found (only 7 SNPs and one indel in the coding region), compared to many fold greater divergence between *rbcS* homeologs in other genera (e.g., *Arabidopsis*, Krebbers *et al.*, 1988; *Triticum*, Sasanuma, 2001; *Gossypium*, Gong *et al.*, 2012). 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 structure and did not result in a change in protein structure or folding. Thus, even with little sequence divergence between the progenitor copies, we detected a maternal influence for the retention and expression of *rbcS-1* copies in the allopolyploids, which suggests that factors other than transcript differences may be acting to facilitate cytonuclear coordination.

2.5.2 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 (Soltis & Soltis, 2000; Wendel, 2000; Chen, 2007b). Homeolog loss is one genetic modification commonly observed post-polyploidization in different species (*Tragopogon*, Koh *et al.*, 2010; *Brassica*, Xiong *et al.*, 2011; Wheat, Feldman *et al.*, 2012; *Arabidopsis*, Matsushita *et al.*, 2012; Cotton, Wendel *et al.*, 2012). However, the homeolog losses observed previously in *Tragopogon* polyploids were species-specific, in that *T. dubius* homeologs were lost more frequently than *T. pratensis* copies in both short- and long-liguled *T. miscellus* and more often than *T. porrifolius* homeologs in *T. mirus* (Tate *et al.*, 2006c; Buggs *et al.*, 2009; Tate *et al.*, 2009a; Koh *et al.*, 2010; Malinska *et al.*, 2010). None of these loci showed maternal bias in their patterns of loss, as we have found here

with *rbcS-1*. This maternally biased loss suggests that not all homeolog losses in *Tragopogon* are random, but may be more conserved and 'directed' for some genes. In polyploids, homeolog losses may be associated with dosage compensation to efficiently maintain gene regulatory mechanisms (Birchler & Veitia, 2010). In the case of cytonuclear coordination involving multi-subunit complexes like Rubisco, loss of the paternal *rbcS* homeolog and retention of the maternal copy may facilitate the regulatory coordination between the rbcS/rbcL subunits. Examination of other cytonuclear complexes would lend insight to this trend of maintaining genomic balance between nuclear homeologs and their cytoplasmic counterparts.

Changes in duplicate gene expression are another consequence of allopolyploidization (Bottley *et al.*, 2006; Chen, 2007b; Buggs *et al.*, 2010b; Flagel & Wendel, 2010), 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 (Chen & Pikaard, 1997; Adams *et al.*, 2003b; Wang *et al.*, 2006; Chaudhary *et al.*, 2009; Grover *et al.*, 2012). Previous studies on *Tragopogon* identified alterations in expression of homeologous loci, but as with the genomic losses discussed above, those expression differences were species-specific (i.e., *T. dubius* loci silenced more often than *T. pratensis*) (Tate *et al.*, 2006c; Buggs *et al.*, 2010b; Koh *et al.*, 2010). In this study, expression of parental *rbcS-1* homeologs was biased toward the maternal parent, presumably to facilitate cytonuclear accommodation with the maternally inherited *rbcL* product. Individuals from short-liguled *T. miscellus* populations (Spokane, Spangle, Oakesdale, Garfield, Rosalia, Moscow and Troy) preferentially expressed the *T. pratensis rbcS-1* homeolog (maternal), while all individuals from the long-liguled population (Pullman) showed preferential expression of the *T. dubius rbcS-1* homeolog (maternal). This finding again indicates repeatability of cytonuclear coordination in reciprocally formed polyploids.

The successful establishment of hybrids and allopolyploids requires coordination between the maternally inherited cytoplasmic (chloroplast and mitochondrial) and the biparentally inherited nuclear genomes to facilitate genomic stability (Fishman & Willis, 2006; Barr & Fishman, 2011). 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 (Levin, 2003). Our study provides the first glimpse as to how long it takes for this coordination to be established. All the synthetic polyploids retained and expressed both parental homeologs, while 19% of the naturally occurring plants of allotetraploid *T. miscellus* analyzed showed an unequal expression pattern that was biased towards the maternal parent. Further, another 13% of the naturally occurring polyploids showed loss of the paternal homeolog as another mechanism to facilitate interaction between the merged genomes. Given that the naturally occurring *Tragopogon miscellus* populations are less than 100 years old (~50 generations as they are biennials), resolution of cytonuclear incompatibilities may only take a few generations to be established. Examination of the synthetic lineages over successive generations would lend valuable insight as to when these changes start to occur. In comparison, both parental *rbcS* homeologs were retained and expressed in *Gossypium* allopolyploids, which are several million years old (Wendel & Cronn, 2003; Gong *et al.*, 2011). *Gossypium* has a longer generation time, so accommodation would proceed more slowly on an absolute temporal scale in *Gossypium* compared to *Tragopogon.* Interestingly, the parental 'A' and 'D' genome copies of *rbcS* also showed evidence of non-reciprocal translocation in the allotetraploids, but not an F_1 hybrid, suggesting that such changes to accommodate cytonuclear interactions may take some time to establish in other groups as well.

The biased retention and expression of maternal *rbcS-1* homeologs in individuals from different populations indicates repeatability to this evolutionary trajectory because each population of *T. miscellus* represents an independent formation (Soltis *et al.*, 1995a;

Soltis & Soltis, 1999; Soltis *et al.*, 2004; Symonds *et al.*, 2010). Although the majority of the populations showed this 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. In a previous study of homeolog loss in *T. miscellus* (Tate *et al.*, 2009a), this population showed a greater number of homeolog losses (individual 2625-3 in particular) than other populations. Although in general there seems to be a recurrent pattern toward maternal bias, some populations may favor different parental combinations to facilitate cytonuclear interactions.

Given that the predicted protein structure of both parental *rbcS-1* homeologs is the same, exactly what has driven differential expression of the maternal copy of *rbcS-1* in the natural 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.* (1989) 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, *Lycopersicon,* 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 polymorphisms between the parents occur there. 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 suggest differential selection on the *rbcS-1* copies under some conditions. Therefore, although the mechanism for cytonuclear coordination is not yet known, clearly the maternal *rbcS-1* homeolog seems to be more compatible with the maternally inherited *rbcL* for successful cytonuclear coordination in *T. miscellus* polyploids. Further research involving protein-protein interactions between rbcS/rbcL subunits in *T. miscellus* would be helpful to further understand the complexities of these cytonuclear interactions.

2.6 Acknowledgements

We thank Muhammad Faisal for assistance in predicting promoter and terminator elements, and Afsana Islam for her guidance in genome walking experiments. 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 Tina Sehrish.

2.7 Supplementary material

Table S2.1 List of natural and synthetic polyploids (*T. miscellus*) and diploid parents (*T. dubius* and *T. pratensis*) examined. Data are summarized from genomic DNA and cDNA sequencing, genomic 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.

Species	Population/ synthetics	Short or long liguled	Unique no./ Cross no.	Collection no. / S_0 line no.	Individual/ S_1 line no.	Retention	Expression level
T. dubius	Spokane			2615	22	D	D
T. dubius	Spokane			2615	4	D	D
T. dubius	Spangle		1987	2616	8	D	D
T. dubius	Spangle		1988	2616	11	D	D
T. dubius	Oakesdale		1980	2670	9	D	D
T. dubius	Oakesdale		1981	2670	10	D	D
T. dubius	Garfield		1938	2687	3	D	D
T. dubius	Garfield		1943	2687	11	D	D
T. dubius	Pullman		1073	2613	1	D	D
T. dubius	Pullman		1095	2613	35	D	D
T. dubius	Troy			2683	1	D	D
T. dubius	Troy			2683	7	D	D
T. pratensis	Spangle		2138	2692	1	P	P
T. pratensis	Spangle		2139	2692	2	P	P
T. pratensis	Oakesdale		2150	2672	4	P	P
T. pratensis	Oakesdale		2151	2672	5	P	P
T. pratensis	Garfield		2136	2689	15	P	P
T. pratensis	Garfield		2137	2689	17	P	P
T. pratensis	Moscow		1049	2608	31	P	P
T. pratensis	Moscow			2608	35X	P	P
T. miscellus	Spokane	Short		2664	3	Both	-
T. miscellus	Spokane	Short		2664	5a	Both	-
T. miscellus	Spokane	Short		2664	5b	Both	-
T. miscellus	Spokane	Short		2664	6	Both	-
T. miscellus	Spokane	Short		2617	1	Both	-
T. miscellus	Spokane	Short		2617	4	Both	-
T. miscellus	Spokane	Short		2617	6	Both	-
T. miscellus	Spokane	Short		2617	7	Both	-
T. miscellus	Spokane	Short		2617	8	Both	-
T. miscellus	Spokane	Short		2617	9	Both	-
T. miscellus	Spokane	Short		2617	12	Both	-
T. miscellus	Spokane	Short		2617	21	Both	-

T. miscellus	Spangle	Short	2121	2693	3	Both	-
T. miscellus	Spangle	Short	2123	2693	5	Both	-
T. miscellus	Spangle	Short	2125	2693	7	Both	P> D
T. miscellus	Spangle	Short	2126	2693	8	Both	Equal expression
T. miscellus	Spangle	Short	2127	2693	9	P	P
T. miscellus	Spangle	Short	2129	2693	11	P	P
T. miscellus	Spangle	Short	2130	2693	12	Both	Equal expression
T. miscellus	Spangle	Short	2131	2693	13	Both	-
T. miscellus	Spangle	Short	2132	2693	14	Both	-
T. miscellus	Spangle	Short	2133	2693	15	Both	-
T. miscellus	Rosalia	Short		2667	3	Both	-
T. miscellus	Rosalia	Short		2667	4	Both	-
T. miscellus	Oakesdale	Short	2055	2671	1	Both	-
T. miscellus	Oakesdale	Short	2056	2671	2	Both	P>D
T. miscellus	Oakesdale	Short	2057	2671	3	Both	-
T. miscellus	Oakesdale	Short	2058	2671	4	Both	Equal expression
T. miscellus	Oakesdale	Short		2671	6	-	Equal expression
T. miscellus	Oakesdale	Short	2061	2671	7	Both	Equal expression
T. miscellus	Oakesdale	Short	2062	2671	8	-	Equal expression
T. miscellus	Oakesdale	Short	2063	2671	9	Both	-
T. miscellus	Oakesdale	Short	2064	2671	10	Both	-
T. miscellus	Oakesdale	Short	2065	2671	11	Both	P>D
T. miscellus	Garfield	Short	2099	2688	1	Both	-
T. miscellus	Garfield	Short	2100	2688	2	Both	Equal expression
T. miscellus	Garfield	Short	2103	2688	5	Both	-
T. miscellus	Garfield	Short	2104	2688	6	Both	Equal expression
T. miscellus	Garfield	Short	2106	2688	8	P	P
T. miscellus	Garfield	Short		2688	10	Both	Equal expression
T. miscellus	Garfield	Short	2109	2688	11	Both	Equal expression
T. miscellus	Garfield	Short	2110	2688	12	Both	P>D
T. miscellus	Garfield	Short	2111	2688	13	Both	-
T. miscellus	Albion	Short		2625	1	Both	-
T. miscellus	Albion	Short		2625	2	Both	-
T. miscellus	Albion	Short		2625	3	D	-
T. miscellus	Albion	Short		2625	5	Both	-
T. miscellus	Albion	Short		2625	6	D	-
T. miscellus	Albion	Short		2625	8	D	-
T. miscellus	Albion	Short		2625	9	Both	-
T. miscellus	Albion	Short		2625	10	Both	-

T. miscellus	Pullman	Long		2605	3	-	Equal expression
T. miscellus	Pullman	Long	1194	2605	4	Both	-
T. miscellus	Pullman	Long		2605	5	-	Equal expression
T. miscellus	Pullman	Long	1196	2605	7	Both	Equal expression
T. miscellus	Pullman	Long		2605	9	-	D>P
T. miscellus	Pullman	Long		2605	10	-	Equal expression
T. miscellus	Pullman	Long	1203	2605	13	Both	D>P
T. miscellus	Pullman	Long	1204	2605	14	Both	Equal expression
T. miscellus	Pullman	Long	1210	2605	24	Both	Equal expression
T. miscellus	Pullman	Long	1211	2605	28	D	D
T. miscellus	Pullman	Long	1212	2605	29	Both	Equal expression
T. miscellus	Pullman	Long		2605	36	-	Equal expression
T. miscellus	Pullman	Long	1216	2605	42	Both	Equal expression
T. miscellus	Pullman	Long	1221	2605	46	Both	D>P
T. miscellus	Moscow	Short	1154	2604	4	Both	-
T. miscellus	Moscow	Short		2604	8	-	Equal expression
T. miscellus	Moscow	Short	1157	2604	10	Both	-
T. miscellus	Moscow	Short	1158	2604	11	Both	Equal expression
T. miscellus	Moscow	Short	1162	2604	15	Both	Equal expression
T. miscellus	Moscow	Short		2604	17	-	P
T. miscellus	Moscow	Short	1171	2604	20	Both	Equal expression
T. miscellus	Moscow	Short		2604	21	-	Equal expression
T. miscellus	Moscow	Short	1173	2604	22	P	P
T. miscellus	Moscow	Short	1176	2604	24	Both	-
T. miscellus	Moscow	Short		2604	26	-	Equal expression
T. miscellus	Moscow	Short		2604	29	-	Equal expression
T. miscellus	Moscow	Short		2604	31	-	Equal expression
T. miscellus	Moscow	Short	1186	2604	35	Both	-
T. miscellus	Moscow	Short	1402	2604	43	Both	P>D
T. miscellus	Moscow	Short		2604	48	-	Equal expression
T. miscellus	Moscow	Short		2604	49	-	Equal expression
T. miscellus	Troy	Short		2682	1	Both	-
T. miscellus	Troy	Short		2682	2	Both	-
T. miscellus	Troy	Short		2682	3	Both	-
T. miscellus	Troy	Short		2682	4	Both	-
T. miscellus	Troy	Short		2682	5	P	-

T. miscellus	Troy	Short		2682	6	Both	-
T. miscellus	Troy	Short		2682	7	Both	-
T. miscellus	Troy	Short		2682	11	Both	-
T. miscellus	Troy	Short		2682	12	Both	-
T. miscellus	Troy	Short		2682	11	Both	-
T. miscellus	Troy	Short		2682	12	Both	-
T. miscellus	S_0	Short	111	1	-	Both	-
T. miscellus	S_1	Short	111	1	7	Both	-
T. miscellus	S_1	Short	111	1	8	Both	-
T. miscellus	S_1	Short	111	1	23	Both	Equal expression
T. miscellus	S_0	Short	111	4	-	Both	-
T. miscellus	S_1	Short	111	4	4	Both	-
T. miscellus	S_1	Short	111	4	17	Both	Equal expression
T. miscellus	S_0	Short	111	5	-	Both	-
T. miscellus	S_1	Short	111	5	7	Both	-
T. miscellus	S_1	Short	111	5	9	Both	-
T. miscellus	S_1	Short	111	5	18	Both	Equal expression
T. miscellus	S_0	Short	111	7	-	Both	-
T. miscellus	S_1	Short	111	7	11	Both	Equal expression
T. miscellus	S_1	Short	67	2	1	Both	-
T. miscellus	S_1	Short	67	2	4	Both	-
T. miscellus	S_1	Short	67	2	7	Both	Equal expression
T. miscellus	S_1	Short	67	3	6	Both	-
T. miscellus	S_1	Short	67	3	8	Both	-
T. miscellus	S_0	Short	79	1	-	Both	-
T. miscellus	S_1	Short	79	1	3	Both	-
T. miscellus	S_1	Short	79	11	3	Both	-
T. miscellus	S_0	Long	119	2	-	Both	-
T. miscellus	S_0	Long	129	7	-	Both	-
T. miscellus	S_1	Long	129	7	1	Both	-
T. miscellus	S_1	Long	129	7	14	Both	Equal expression

Table S2.2 Transcription factor binding sites in *rbcS-1* promoter region as determined by Plant Promoter Analysis Navigator (PlantPAN).

Transcription factor	Site	Sequence	Species
AGL3	11	atacCCATAacgtatcag	*Arabidopsis*
Athb-1	254	tgaggATTATatca	*Arabidopsis*
ATHB-9	91	ttggttaATGATaaggcat	*Arabidopsis*
CDC5	22	gtaTCAGCctc	*Arabidopsis*
PIF3	65	ccctgcCACGTgtcacat	*Arabidopsis*
ABRELATED1	72	ACGTG	*Arabidopsis*
ANAERO1CONSENSUS	202	TTTGTTT	Maize/*Arabidopsis*/Pea/Barley/Rice
CACGTGMOTIF	71	CACGTG	Tomato/*Arabidopsis*/Snapdragon/Wheat
Core	1	ATTA	*Arabidopsis*
Core	108	ATTA	*Arabidopsis*
GATABOX	100	GATA	*Petunia/Arabidopsis/*Rice
GATABOX	196	GATA	*Petunia/Arabidopsis/*Rice
GATABOX	263	TATC	*Petunia/Arabidopsis/*Rice
GBOXLERBCS	69	GCCACGTGT	Tomato*/Arabidopsis*
GT1CONSENSUS	108	ATTACC	Pea/Oat/Rice/Tobacco/*Arabidopsis*
HY5AT	68	TGCCACGTGTCA	*Arabidopsis*
IBOXCORENT	100	GATAAGG	Tobacco
IBOXCORENT	48	GATAAGG	tobacco
IBOX	100	GATAAG	Tomato/*Arabidopsis*
IBOX	48	GATAAG	Tomato/*Arabidopsis*
LRENPCABE	68	TGCCACGT	Tobacco
MYB2COREATCYCB1	213	CCGTT	*Arabidopsis*
RBCSCONSENSUS	272	TTGGATT	Tomato/*petunia*/Tobacco/Pea
SORLIP1AT	158	GCCAC	*Arabidopsis*
SORLIP1AT	69	GCCAC	*Arabidopsis*
WBOXNTERF3	82	TGACC	Tobacco

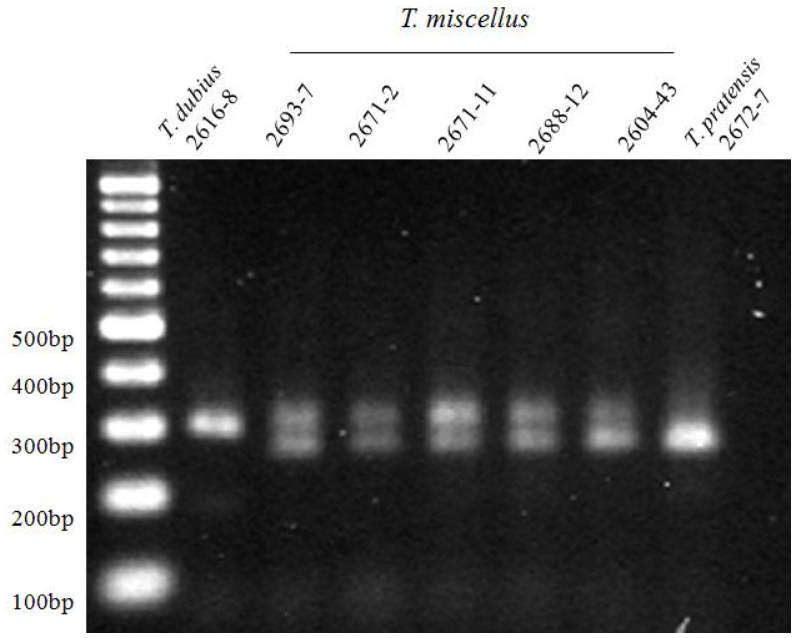
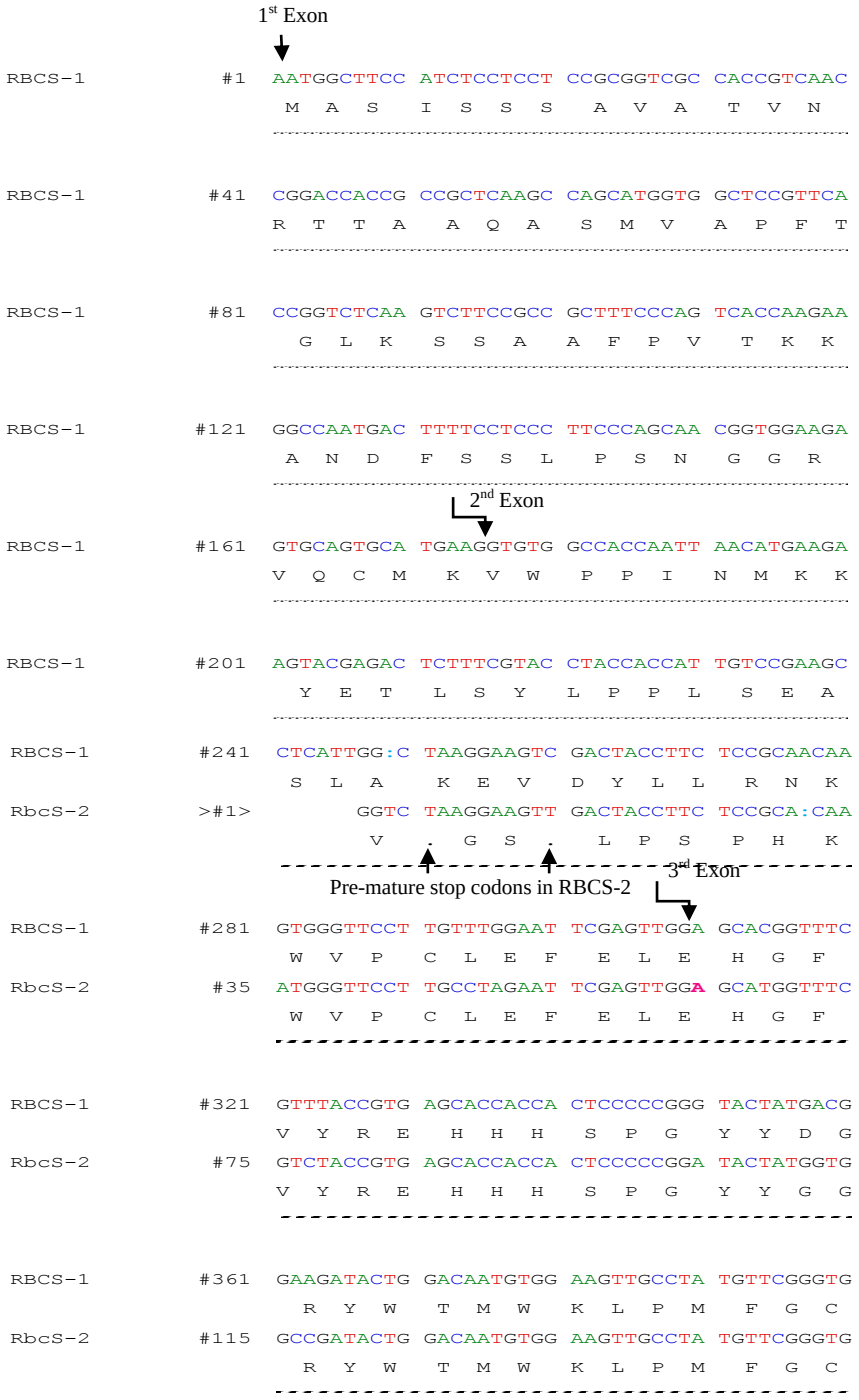
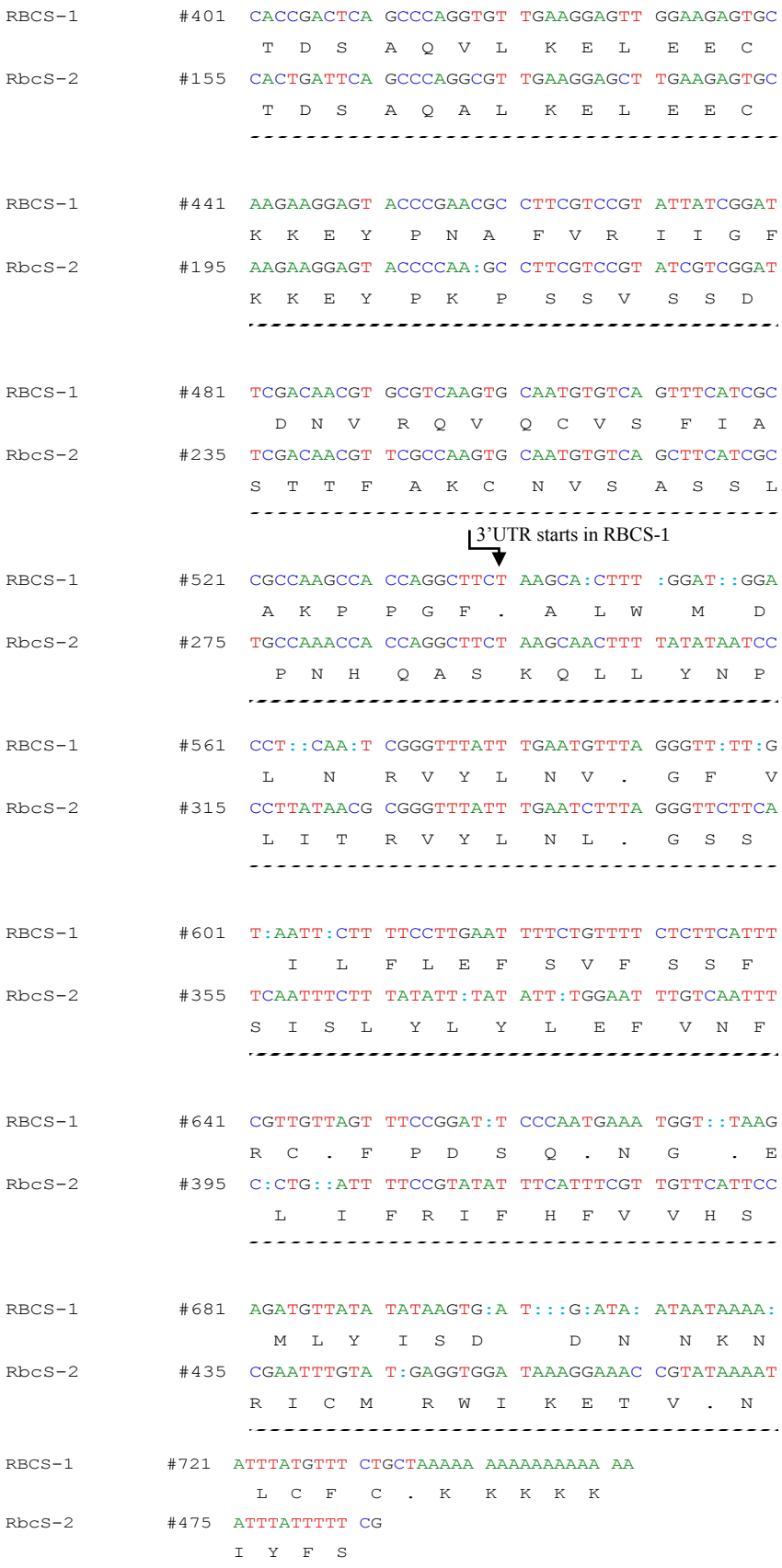


Fig. S2.1 cDNA-CAPS for TDF-85 as a control to check equal expression of parental copies in *T. miscellus* polyploids showing biased maternal expression for *rbcS-1*.

Fig. S2.2 Alignment of *rbcS-1* and *rbcS-2* cDNA sequences of *Tragopogon dubius* along with protein translation for both genes. Arrows indicate position of exons and premature stop codons (in *rbcS-2*).




RBCS-1 #401 CACCGACTCA GCCCAGGTGT TGAAGGAGTT GGAAGAGTGC
T D S A Q V L K E L E E C
RbcS-2 #155 CACTGATTCA GCCCAGGCGT TGAAGGAGCT TGAAGAGTGC
T D S A Q A L K E L E E C
RBCS-1 #441 AAGAAGGAGT ACCCGAACGC CTTCGTCCGT ATTATCGGAT
K K E Y P N A F V R I I G F
RbcS-2 #195 AAGAAGGAGT ACCCCAA:GC CTTCGTCCGT ATCGTCGGAT
K K E Y P K P S S V S S D
RBCS-1 #481 TCGACAACGT GCGTCAAGTG CAATGTGTCA GTTTCATCGC
D N V R Q V Q C V S F I A
RbcS-2 #235 TCGACAACGT TCGCCAAGTG CAATGTGTCA GCTTCATCGC
S T T F A K C N V S A S S L
3'UTR starts in RBCS-1
RBCS-1 #521 CGCCAAGCCA CCAGGCTTCT AAGCA:CTTT :GGAT::GGA
A K P P G F . A L W M D
RbcS-2 #275 TGCCAAACCA CCAGGCTTCT AAGCAACTTT TATATAATCC
P N H Q A S K Q L L Y N P
RBCS-1 #561 CCT::CAA:T CGGGTTTATT TGAATGTTTA GGGTT:TT:G
L N R V Y L N V . G F V
RbcS-2 #315 CCTTATAACG CGGGTTTATT TGAATCTTTA GGGTTCTTCA
L I T R V Y L N L . G S S
RBCS-1 #601 T:AATT:CTT TTCCTTGAAT TTTCTGTTTT CTCTTCATTT
I L F L E F S V F S S F
RbcS-2 #355 TCAATTTCTT TATATT:TAT ATT:TGGAAT TTGTCAATTT
S I S L Y L Y L E F V N F
RBCS-1 #641 CGTTGTTAGT TTCCGGAT:T CCCAATGAAA TGGT::TAAG
R C . F P D S Q . N G . E
RbcS-2 #395 C:CTG::ATT TTCCGTATAT TTCATTTCGT TGTTCATTCC
L I F R I F H F V V H S
RBCS-1 #681 AGATGTTATA TATAAGTG:A T:::G:ATA: ATAATAAAA:
M L Y I S D D N N K N
RbcS-2 #435 CGAATTTGTA T:GAGGTGGA TAAAGGAAAC CGTATAAAAT
R I C M R W I K E T V . N
RBCS-1 #721 ATTTATGTTT CTGCTAAAAA AAAAAAAAAA AA
L C F C . K K K K K
RbcS-2 #475 ATTTATTTTT CG
I Y F S

CHAPTER THREE

This chapter is in preparation as:

***Sehrish T**, Tate JA, Soltis DE, Soltis PS, and Symonds VV . Genetic characterization of synthetic Tragopogon polyploids using microsatellite markers. In prep. for BMC Plant Biology.*

3. Genetic characterization of synthetic *Tragopogon* polyploids using microsatellite markers

3.1 Abstract

Polyploidy brings about extensive genetic and epigenetic modifications. Synthetic polyploids have proven to be a dynamic and useful resource to study the speed of these genetic changes following whole genome duplication compared to natural polyploids. In the present study, we studied the genetic structure of synthetic polyploid lineages of *Tragopogon mirus* and *T. miscellus* in the context of their relationship with each other. Use of 13 progenitor-specific microsatellite markers resulted in the analysis of 17 loci from 73 synthetic polyploid lineages of *T. mirus* and *T. miscellus* included in this study along with their diploid parents. Neighbor-net analyses distinguish clusters of these polyploid lineages along with their parents based on genetic distance. Although genotypes other than those discovered in natural populations were observed, three clusters roughly corresponding to the multilocus genotypes discovered at high frequencies in the natural *Tragopogon* polyploids were also observed in the synthetics. The data show that significant variation has been contributed to the synthetic polyploids by their parents. Interestingly, synthetic lineages which appeared to be least successful beyond the F_1 and S_0 generations deviated the most from the multilocus genotypes observed in natural populations of *T. mirus* and *T. miscellus*, suggesting a potential genetic barrier to successful polyploidy establishment. Present analysis of genetic structure of synthetic lineages would provide a strong comparison for future studies on the consequences of polyploidy in these synthetic lineages.

3.2 Introduction

Although polyploidy was first described more than a century ago (Lutz 1909), its prevalence in plant evolution has only become clear since last few decads. With many examples of both ancient and modern polyploidy events inferred from plant whole genome sequences, polyploidy has emerged as a leading force in plant evolution and diversification (Soltis & Soltis, 1999; Soltis *et al.*, 2009). Indeed, it now appears that all seed plants share a whole genome duplication event and all angiosperms share another. Additionally, more recent polyploidy events provide a continuum from ancient events through to modern polyploids that have arisen in just the last 200 years. Such a temporal sequence of whole genome duplication (WGD) events provides a fantastic opportunity to examine the consequences of polyploidy at several scales.

Unsurprisingly, the documentation of polyploidy's prevalence throughout plant evolution has fueled a broad resurgence in the study and interest in whole genome duplication and the fates of individual gene duplicates. "What are the consequences of polypoidy?" is now being asked at many levels. The answers thus far include that polyploidy can lead to genome instability, chromosomal rearrangement, translocations, gene loss, irregular sequence evolution of duplicated genes, and epigenetic changes like DNA methylation and transposon activation (Chen, 2007b; Chester *et al.*, 2011; Yaakov & Kashkush, 2011). Many of these evolutionary changes take place immediately after polyploid formation, while others occur gradually over successive generations (Mestiri *et al.*, 2010; Yang *et al.*, 2011a).

Identifying the events that immediately follow polyploid formation and how those events might affect the evolutionary outcomes of polyploidy remain important goals. Though there are now many functionally polyploid species recognized, relatively few are known to be of recent origin. Indeed, those that are very recent provide many novel insights (*Tragopogon*

miscellus and *Tragopogon mirus*, Ownbey, 1950; *Senecio cambrensis*, Ashton & Abbott, 1992; *Spartina anglica*, Baumel *et al.*, 2001). A complementary approach to working on recent naturally formed polyploids is the generation of synthetic polyploids, which afford the opportunity to examine the immediate effects of polyploidy and to track changes through generations from inception. Synthetic polyploids mirror natural plant systems and are excellent models for examining the immediate outcomes of polyploidization (Song *et al.*, 1995; Adams *et al.*, 2003b; Tate *et al.*, 2009b). While even work on recently formed polyploids can be complicated by a lack of knowledge around the generation age of a group, synthetic polyploids provide an opportunity to track changes on a known generation scale. Moreover, recurrent formation, which appears to be a common feature of natural polyploid species, and subsequent gene flow can present obstacles to definitively identifying independent origins (Soltis & Soltis, 1995; 1999). Further, the progenitors of natural polyploids are not always readily identifiable because much time may have elapsed since their formation which results in either extinction of parents or extensive changes in parental lineages, for instance *Gossypium* (Wendel & Cronn, 2003) and *Nicotiana* (Clarkson *et al.*, 2004), whereas for synthetic polyploids, their actual parents and genotypes are known (Song *et al.*, 1995; Malinska *et al.*, 2010).

Recently, a number of synthetic polyploids have been produced to study genetic and epigenetic modifications following WGD. For example, loss of parental genomic sequences was demonstrated in synthetic wheat (*Triticum* spp.) allopolyploids (Feldman *et al.*, 1997; Liu *et al.*, 1998; Ma & Gustafson, 2008; Tang *et al.*, 2012), both instantaneously following formation of the polyploid (synthetic allotetraploids, Shaked *et al.*, 2001; Kashkush *et al.*, 2002b) and in subsequent generations (synthetic allohexaploids, Ma *et al.*, 2004). Similarly, meiotic irregularities leading to aneuploidy (Mestiri *et al.*, 2010) and epigenetic remodeling involving changes in the DNA methylation pattern were observed in genomic regions flanking

a retroelement *Veju* across S_1 through S_5 synthetic generations of wheat polyploids (Kraitshtein *et al.*, 2010). Likewise, synthetic polyploids of *Brassica* were used to study the genetic basis of flowering time variation in four homeologs of *FLOWERING LOCUS C (FLC)*(Pires *et al.*, 2004b). Extensive DNA methylation and loss and/or gain of novel restriction fragments was observed in early generations of *Brassica napus* synthetic polyploids via RFLP mapping (Song *et al.*, 1995; Lukens *et al.*, 2006); later this loss and gain of parental fragments was found to be due to homeologous non-reciprocal transpositions (HNRTs) between parental chromosomes A1 and C1 (Gaeta *et al.*, 2007). Similarly, synthetic cotton (*Gossypium* spp.) polyploids have shown organ-specific silencing of duplicated genes (Adams *et al.*, 2004) and unequal contribution of homeologous genes to the transcriptome (Adams *et al.*, 2003b; Flagel & Wendel, 2010; Dong & Adams, 2011b). Synthetic polyploids of *Arabidopsis* have shown phenotypic instability and gene silencing caused by DNA methylation (Comai *et al.*, 2000; Madlung *et al.*, 2002; Madlung *et al.*, 2005; Wright *et al.*, 2009; Matsushita *et al.*, 2012) and work on *Nicotiana* synthetic polyploids has provided evidence for concerted evolution of rDNA repeats, preferential elimination of paternal genomic sequences in S_4 generation polyploids, and various morphological changes post-polyploidization (Kovarik *et al.*, 2004; Skalicka *et al.*, 2005; Anssour *et al.*, 2009; Anssour & Baldwin, 2010). In each case, synthetic polyploids have proven to be valuable resources for the analysis of the early stages of polyploid genome evolution.

Among the best examples of recent and recurrent polyploid formation are the allopolyploid species, *Tragopogon mirus* and *T. miscellus.* The genus is comprised of ~150 species spanning Eurasia (Soltis *et al.*, 2004). In the early 1900s, three diploid ($2n$=$2x$ =12) *Tragopogon* species (*T. dubius*, *T. pratensis*, and *T. porrifolius*) were introduced to eastern Washington State and neighboring Idaho (USA), an area known as the Palouse Region. Since their introduction, these species have produced two allopolyploid ($2n$ = $4x$ = 24) species: *T.*

mirus resulting from *T. dubius* × *T. porrifolius* and *T. miscellus* produced by *T. dubius* × *T. pratensis* (Ownbey, 1950; Ownbey & McCollum, 1954; Soltis & Soltis, 1995; 1999). Natural polyploids of *Tragopogon mirus* and *T. miscellus* have formed recurrently in the past 80 years, (reviewed in Soltis *et al.*, 2004). In addition to morphological and cytological studies by Ownbey and McCollum, (1953; 1954), the number of independent origins has been estimated several times in the past using biochemical or molecular approaches (Soltis & Soltis, 1989; Soltis & Soltis, 1991; Soltis & Soltis, 1995; Symonds *et al.*, 2010). *Tragopogon miscellus* has been suggested to have originated independently as many as 21 times and *T. mirus* has formed at least 11 times in the Palouse region (Soltis & Soltis, 2000).

In the most recent population genetic analysis of *T. mirus* and *T. miscellus*, Symonds *et al*. (2010) generated progenitor-specific microsatellite markers from the parental diploid species (*T. dubius, T. pratensis* and *T. porrifolius*) to analyze recurrent formations and to identify the genetic contributions of the diploid progenitors to each polyploid species. Among the most interesting findings was that the common diploid progenitor, *T. dubius*, has contributed relatively few multilocus genotypes to both *T. miscellus* and *T. mirus*. The three common genotypes (Types I-III) observed in both polyploid species are consistently prescribed by five *T. dubius* microsatellite markers. As the genomic positions of these loci are unknown, it is possible that some markers may be pysically linked; however, the pattern often extends to all 10 T. dubius loci, arguing for an alternative explanation. It was hypothesized that the three common genotypes represent either (1) the only genotypes present in *T. dubius* during the time(s) of polyploid formation or (2) the only genotypes able to successfully produce F_1s or polyploids in nature.

Early attempts at resynthesizing *Tragopogon mirus* and *T. miscellus* polyploids were made by Ownbey and McCollum (1953). They generated diploid F_1 hybrids by hand-crossing diploid species, but never saw the generation of spontaneous allopolyploid plants. More than

half a century later, the process was helped along by the application of colchicine to interspecific F_1s generated by hand-crossing to produce several lineages of synthetic allopolyploid *Tragopogon mirus* and *T. miscellus* (Tate *et al.*, 2009b). Several hundred hand-crosses yielded 389 offspring that were treated with colchicine and subsequently screened to identify those of mixed parentage and that were polyploid (Tate *et al.*, 2009b). These synthetic polyploids are revealing insights into the early generation genetic structure of natural polyploid systems. Lim *et al*. (2008) found rapid changes following polyploid formation in S_0 and S_1 generations of *T. mirus* including translocations, chromosomal abnormalities and variable expression patterns of ribosomal DNA (rDNA) loci. Similarly, meiotic irregularities involving multivalent formation and bridges were commonly observed in the S_0 generation of *T. mirus* and *T. miscellus* (Tate *et al.*, 2009b). However, Buggs *et al*. (2009) found no evidence of homeolog losses in the S_1 generation of *T. miscellus*. Recently, inheritance of parental rDNA loci was studied in synthetic polyploids from S_0, S_1, and S_2 generations of *T. mirus* and *T. miscellus*. There was considerable deviation from copy number additivity with reduction of rDNA loci from *T. dubius* compared to *T. porrifolius* and *T. pratensis* (Malinska *et al.*, 2010; Malinska *et al.*, 2011).

As independent synthetic lineages of any one polyploid may reveal different outcomes with regard to genomic rearrangements, gene silencing, etc., it is difficult to know whether such differences reflect an element of stochasticity in the genomic processes following WGD or if there might be a genetic basis to such varied results. This appears to be an underexplored feature of synthetic polyploid research (although see Song *et al.*, 1995; Shaked *et al.*, 2001; Madlung *et al.*, 2005; Skalicka *et al.*, 2005) and one that may provide greater comparative value to synthetic polyploid research. For the *Tragopogon* synthetic polyploids, although the population origins of each lineage are known, the genetic composition of each lineage, relative to one another and to natural *T. mirus* and *T. miscelllus* has been completely

unknown. In the present study, we genetically characterize the synthetic polyploids of *Tragopogon mirus* and *T. miscellus* from Tate *et al*. (2009b) using the microsatellite markers used by Symonds *et al*. (2010) to analyze the natural polyploids. By combining these resources, we are able to address the following questions: (1) What are the contributions of the diploid progenitors to genetic variation in synthetic polyploids (*T. mirus and T. miscellus*)? (2) How much genetic variation is captured by the synthetic lineages relative to natural *T. mirus* and *T. miscellus*? (3) Do the multi-locus *T. dubius* genotypes observed in the natural polyploids also predominate in the synthetics? (4) Do patterns of synthetic *Tragopogon* polyploid formation and persistence (through generations) support the hypothesis that only particular genotypes can form diploid hybrids or polyploids?

3.3 Materials and Methods

3.3.1 Plant material

The current study utilized both synthetic polyploids (F_1 diploid hybrid, S_0 and S_1 generation) and their diploid parents. The plant materials used in this study are described in Tate *et al*. (2009b). Figure 3.1 depicts how the synthetic lineages were generated.

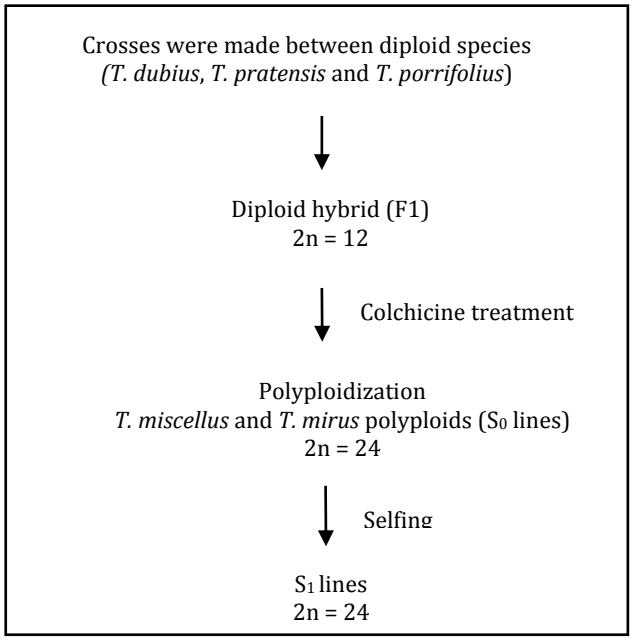


Fig. 3.1 Formation of synthetic polyploids.

For diploid parents and synthetic polyploids (F_1 hybrids and S_0 generation) plant materials were the same as those used by Tate *et al*. (2009b). For the later synthetic polyploid generation (S_1), leaf tissue was collected from 4 week old seedlings grown under standard conditions (23 ± 2°C day / 18°C night and 16 hours day: 8 hours night). DNA was extracted from fresh or silica gel-dried leaf tissue after homogenizing it in a bead mill (MagNA Lyser, Roche); DNA was extracted using a modified CTAB protocol (Doyle & Doyle, 1987). Where possible, those F_1 hybrids that did not survive beyond the S_0 generation or died after germination or after colchicine treatment were also included in this study. Also, where possible, the exact diploid progenitors of the synthetic polyploids were included for comparative analysis. The synthetic polyploids (S_0 and S_1 generation) and diploid F_1 hybrids examined in this study along with their exact parental lineages are listed in Table 3.1. Details about the number of successful S_0 and S_1 lineages produced can be found in Tate *et al*. (2009b).

3.3.2 Microsatellite PCR

Microsatellite markers described by Symonds *et al.* (2010) were amplified from individuals using 13 primer pairs. These markers were designed specifically from *Tragopogon dubius* and *T. porrifolius. T. dubius* is a common parent for *T. miscellus* and *T. mirus* polyploids, so microsatellite markers designed from *T. dubius* amplify from both polyploid species. Markers designed from *T. porrifolius* amplify from *T. pratensis* as well so provide markers for the alternative (non-*T. dubius*) genome in both polyploids. Some primer pairs amplify more than one locus, one from each progenitor species. Detailed information on each microsatellite locus can be found in Symonds *et al*. (2010). Microsatellite loci were amplified using a tailed protocol based on Schuelke (2000). Each amplification was prepared in a total volume of 10µl, using 1x PCR buffer from New England Biolabs, U.K. (NEB), 50 µM each dNTPs, 4.5 µM of a dye labeled M13 tailed primer and reverse primer, 0.2 µM of tailed forward primer, 0.5 Unit of Taq

polymerase (NEB) and ~50ng of genomic DNA. M13 tail primers were labeled with FAM, VIC or NED dyes. The PCR profile is as follows: denaturation at 94°C for 3 min followed by 30 cycles of 94°C for 30 sec, 52°C for 30 sec, 72°C for 45 sec, with a final extension at 72°C for 20 min. After PCR, three markers each labeled with a different dye (FAM, VIC or NED) were pooled in equal volumes and 2 µl of this mixture were added into sample loading solution and fluorescently labeled CASS size standard (Symonds & Lloyd, 2004). Fragments were separated on an ABI 3730 DNA Analyzer and the results were analyzed using GeneMapper software version 3.7 (ABI).

3.3.3 Comparison of F_1, S_0 and S_1synthetics with natural *Tragopogon* polyploids

To compare the results from the synthetics to the previous results on natural *T. miscellus* and *T. mirus* populations (which were analyzed on a different platform), representative samples from Symonds *et al.* (2010) were included in the current study. Allele sizes from the ABI (this study) were compared with allele sizes from the Beckman-Coulter CEQ DNA analyzer (previous study) for the same samples and a conversion was established based on these comparisons. Allele sizes were consistently found to differ by one, two, or three base pairs, depending on the marker. For example, the *T. dubius* marker D1048 amplified alleles of 214 bp and 222 bp on the CEQ DNA analyzer, while their size on the ABI DNA analyzer was 217 bp and 225 bp, respectively. This information was used to generate a conversion for each locus. This allowed for the direct comparison of new samples to the Symonds et al. (2010) data.

Moreover, for an accurate assessment of genetic variation among synthetic polyploids and to identify the genetic structure of synthetic polyploids compared to the naturals, three generic genotypes covering the 10 *T. dubius* loci were generated to represent those multilocus genotypes (I-III) found in the natural polyploid populations; these simplified analyses as each Type could be represented by a single genotype/individual in analyses.

Table 3.1 Crossing information and number of F_1, S_0 and S_1 synthetic polyploid lines examined in the study**.**

Speceis	Lines	Maternal ID	Maternal lineage	Maternal population	Paternal ID	Paternal lineage	Paternal population	No. F_1 lineages included	No. S_0 lineages included	No. S_1 lineages included
T. miscellus– long liguled	119	1103	2615-6*	Spokane	1043	2608-19*	Moscow	1	-	-
T. miscellus– long liguled	75	1113	2615-46	Spokane	1063	2609-17*	Spangle	2	-	-
T. miscellus– long liguled	129	1093	2613-24*	Pullman	1038	2608-1*	Moscow	-	1	7
T. miscellus- short liguled	111	1038	2608-1*	Moscow	1093	2613-24*	Pullman	-	4	24
T. miscellus- short liguled	67	1055	2609-3	Spangle	1296	2616-12	Spangle	-	1	13
T. miscellus- short liguled	63	1063	2609-17*	Spangle	1280	2616-4*	Spangle	7	-	-
T. miscellus- short liguled	79	1057	2609-10*	Spangle	1297	2616-12	Spangle	-	2	7
T. mirus	54	1362	2611-8	Pullman	1075	2613-5	Pullman	-	1	-
T. mirus	70	1007	2611-6	Pullman	1090	2613-23	Pullman	-	1	-
T. mirus	73	1012	2611-11	Pullman	1092	2613-24	Pullman	-	1	5
T. mirus	77	1020	2607-2	Troy	1075	2613-5	Pullman	-	1	3
T. mirus	84	1078	2613-11*	Pullman	1010	2611-8*	Pullman	-	1	1
T. mirus	98	1003	2611-2	Pullman	1079	2613-11*	Pullman	-	1	1
T. mirus	99	1006	2611-3*	Pullman	1079	2613-11*	Pullman	-	1	1
T. mirus	108	1113	2615-46	Spokane	1029	2607-18*	Troy	-	-	1
T. mirus	116	1020	2607-2	Troy	1108	2615-21*	Spokane	-	-	6
T. mirus	121	1110	2615-28	Spokane	1035	2607-22B	Troy	-	-	1
T. mirus	132	1090	2613-23	Pullman	1035	2607-22B	Troy	-	-	2
T. mirus	134	1032	2607-21*	Troy	1093	2613-24*	Pullman	-	1	2
T. mirus	138		2616-1	Spangle	1032	2607-21*	Troy	-	-	1

Note: Parental lineages with an asterisk (*) were included in the analyses.

3.3.4 Microsatellite data analysis

All data were formatted for GenAlEx Version 6.41 (Peakall & Smouse, 2006) and various data partitions were analyzed as below. Genetic distance (GD) and Principal Coordinates Analysis (PCA) via covariance matrix with data standardization was estimated in GenAlEx 6.41. Data were exported from GenAlEx 6.41 to Microsatellite Analyser (MSA) version 4.05 (Dieringer & Schlotterer, 2003) and output files were saved in the tab-delimited format. An individual pair-wise genetic distance matrix was generated using the proportion of shared alleles model in MSA. This output file from MSA was used as the input file for SplitsTree4 (version 4.11.3) (Huson & Bryant, 2006). Neighbor-nets were generated from the input distance matrix file in SplitsTree4.

Several independent neighbor-nets were produced from the available data using the synthetic polyploids along with their known diploid parents. Neighbor-nets with only the *T. dubius* parent along with synthetic polyploid lineages *T. miscellus* and *T. mirus* were based on only *T. dubius* loci. Similarly, separate Neighbor-nets with the other progenitor (*T. pratensis* with *T. miscellus* synthetics and *T. porrifolius* along with *T. mirus* synthetic lineages) were based only on the loci amplified from those parents. Independent neighbor-nets for *T. miscellus* and *T. mirus* based on all loci amplified from their parents were also produced. A *T. porrifolius* marker P1111 was excluded from the analysis in which *T. mirus* was included because marker P1111 gave overlapping allele sizes for the parental species and alleles amplified could not be assigned to one of the parental species or the other (*T. dubius* and *T. porrifolius*). Marker P1111 was also removed from those generic types (I, II and III) that were included in *T. mirus* analyses. So, generic types included in the *T. mirus* analyses were based on 9 *T. dubius* loci while those included in *T. miscellus* analysis were based on 10 *T. dubius* loci.

3.3.5 Exploration of multilocus genotypes in the synthetic polyploids

All data produced from the ABI 3730 DNA Analyzer were scored in GeneMapper version 3.7. To identify synthetic lineages that may have the same multilocus genotypes as those discovered in the natural *Tragopogon* polyploids, data were sorted according to size based on one of the five *T. dubius* loci that were diagnostic for Types I-III. Those individuals matching one of the three Types at the first locus were saved and the remainder discarded. The individuals that were saved from the first pass were then sorted according to the second diagnostic marker and those again matching of the three Type patterns were saved. This continued until only individuals with genotypes matching a Type at either all five or ultimately all 10 *T. dubius* loci were isolated. The numbers of synthetic polyploid lines falling into those multilocus types are listed in Table 3.2.

Table 3.2 Occurrence of the multilocus genotypes in the synthetic polyploids.

	T. mirus		*T.miscellus*	
Genotype	No. of crosses	No. of lineages	No. of crosses	No. of lineages
Type I	4	17	1	16
Type II	6	12	1	1
Type III	0	0	3	28
Others	3	4	2	7

3.4 Results

3.4.1 Amplification efficiency and diversity of microsatellite markers

Microsatellite markers produced specifically from *Tragopogon dubius* and *T. porrifolius* were amplified efficiently from the F_1 hybrids, S_0 and S_1 generations of the two synthetic polyploids (*T. miscellus and T. mirus*) and their diploid progenitors (*Tragopogon dubius, T. porrifolius* and *T. pratensis*) (Tate *et al.*, 2009b). Because *T. dubius* is the common parent of both polyploid species, markers developed from it amplified well from both polyploid species. Markers designed from *T. porrifolius* amplified loci from both *T. porrifolius* and *T. pratensis* and four *T. porrifolius* markers amplified from *T. dubius* as well. Out of seven *T. porrifolius* markers, five amplified two different loci from *T. porrifolius* and *T. pratensis*, while marker P1119-p amplified a *T. porrifolius* locus only from *T. mirus* synthetic polyploids. Marker P1111 amplified similar allele sizes for *T. dubius* and *T. porrifolius* from *T. mirus* polyploids*;* because those alleles could not be assigned to one of the parental species unambiguously, marker P1111 was excluded from all the *T. mirus* analyses.

Three of the seven *T. dubius* individuals (used for making synthetic polyploids) included in our study were heterozygous at some microsatellite loci. Two of those *T. dubius* individuals were from the Pullman, Washington population: individual 1083 (2613-19) was heterozygous at D1054, D1056, D1072-d and P1111-d loci, while individual 1093 (2613-24) was heterozygous at D1025, D1055, P1110-d and P1111-d loci (in the locus name, 'P' and 'D' denote the progenitors *T. porrifolius* and *T. dubius,* respectively from which that marker is designed, while a small letter at the end ('p' and 'd') denotes *T. porrifolius/T. pratensis* and *T. dubius* loci/homeologs, respectively). The third *T. dubius* individual 1103 (2615-6) from the Spokane, Washington population was heterozygous only at the P1111 locus. This heterozygosity of *T. dubius* individuals was reflected in the synthetic polyploids that they parented. For example, *T. dubius* 1093 (2613-24) produced heterozygous alleles of 252bp and

254bp for locus D1025 and in its polyploid derivatives (lines 111, 129 and 134), both alleles were found.

3.4.2 Occurrence of multilocus genotypes in the synthetic polyploids

Analysis of the progenitor-specific microsatellite markers was helpful to determine the frequency of the multilocus genotypes described from natural *T. mirus* and *T. miscellus* by Symonds *et al.* (2010) in the synthetic polyploid lineages. Results for the occurrence of the multilocus genotypes based on the five diagnostic *T. dubius* markers (D1025, D1056, P1110-d, P1112-d and P1119-d) in the synthetic polyploids are listed in Table 3.2. Approximate results are listed in Table 3.2 because some lines align with the Types but had missing data and some lines do not match at all five *T. dubius* loci. The results indicate that Type II was the most common among the crosses (crosses are considered instead of lineages because all the lineages (from S_0 and S_1 generation) produced from a single cross will inherit the same genotype, except where progenitors are heterozygous). This finding is unsurprising when considering that most of the *T. dubius* individuals used to produce synthetic polyploids had a Type II genotype. Out of the seven *T. dubius* parents included in this study, four had Type II, two had Type I and one had Type III.

Of the 20 synthetic polyploid crosses studied (13 from *T. mirus* and 7 from *T. miscellus*), multilocus genotype I was present in five lines (four from *T. mirus* and one from *T. miscellus*), Type II was present in seven lines (six from *T. mirus* and one from *T. miscellus*) and Type III was present in three *T. miscellus* lines. Five lines (three from *T. mirus* and two from *T. miscellus*) did not fit into any of the three Types based on the five *T. dubius* diagnostic loci. These lines had a mix of alleles from the three Types for most of the markers except D1025 which produced novel alleles of 254bp and 260bp (Table 3.2).

3.4.3 Genetic structure of the synthetic polyploids

In order to determine the genetic relationships of synthetic polyploid lineages with each other and to identify the genetic contributions from their diploid parents, neighbor-nets were constructed based on genetic distances derived from microsatellite data.

3.4.3.1 *T. miscellus*

For *T. miscellus*, two neighbor-nets were constructed: one based on all 17 loci from both progenitors and including only the *T. miscellus* synthetics and another based on just *T. dubius* loci; the latter included synthetic *T. miscellus*, the available *T. dubius* parents of the synthetics, and three "generic genotypes" that represent the common multilocus genotypes described from natural *T. miscellus* and *T. mirus* (Figure 3.2). These two neighbor-nets are very similar in structure and are considered together here. All synthetic polyploid lineages (F_1, S_0 and S_1 generations) produced from a single cross typically clustered together. Mainly there were three genotypic clusters that roughly correspond to the three multilocus genotypes (I, II and III). One of the splits separates lineages of crosses 111 and 129 from the other crosses in the neighbor-net; 111 and 129 represent reciprocal crosses from the same parents. Variation within this cluster is due to segregation of alleles at loci for which the *T. dubius* parent was heterozygous (D1025 and P1110-d). These lineages are similar to the Type I genotype except at marker D1112. Lineages 111-1 and 111-7 fell in between the two primary subclusters in this group presumably due to missing data for marker P1110-d (Fig. 3.2a). There was only one S_0 lineage, 119-2, from cross 119. This lineage clustered with the generic Type II genotype and together these were separated from all other individuals by a well-supported split; the 119-2 lineage matches the original multilocus Type II at 8 of 10 *T. dubius* loci (Fig. 3.2b). The remainder of the lineages from crosses 63, 67 and 79 grouped together into one large cluster. These three crosses were made from parents from the same

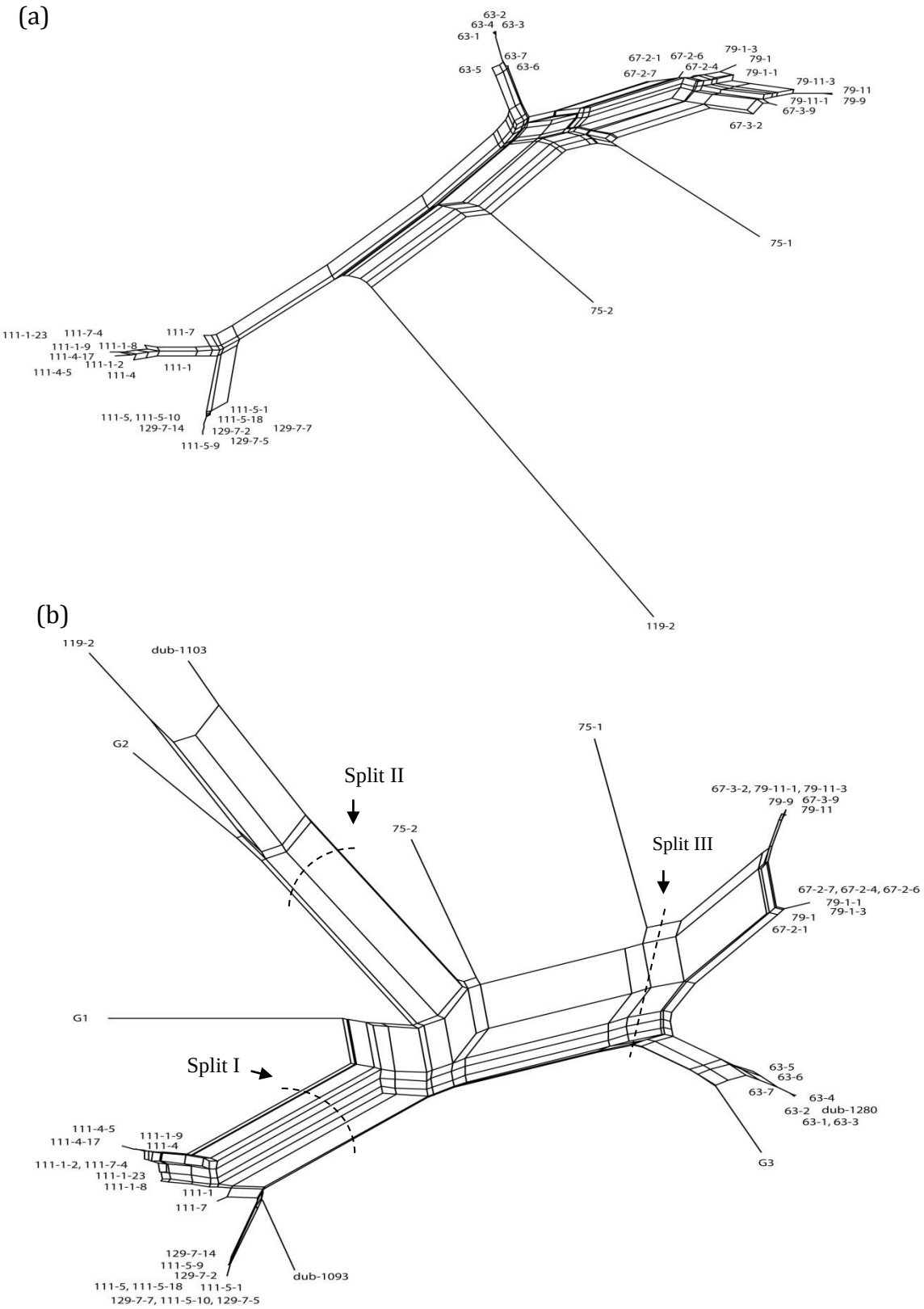


Fig. 3. 2 Neighbor-net of *T. miscellus* synthetic polyploids. (a) Based on all microsatellite loci. (b) Based on only *T. dubius* loci and includes their *T. dubius* parents and generic types named as G1, G2 and G3 based on multilocus genotype I, II and III. Each split is corresponding to the cluster of individuals having distinct multilocus genotype found in the natural polyploid populations.

populations (i.e., the *T. dubius* parents were from the same population and the *T. pratensis* parents were from the same population). In this cluster, lineages of cross 63 were grouped with generic type III; all the lineages produced from this cross match the original type III at 9 *T. dubius* loci (Fig. 3.2b). Lineages from the other two crosses (67 and 79) in this cluster partially match the Type III genotype but differ at locus P1112-d.

Another neighbor-net was generated for synthetic *T. miscellus* and the available *T. pratensis* parents based on 6 loci amplified (Fig. 3.4a). This neighbor-net is divided into two main clusters. One long split separates all the lineages derived from crosses 111, 119 and 129 (cluster A) from other lineages in the neighbor-net based on variation at P1108-p among *T. pratensis* individuals parenting these lines. Three S_1 lineages 67-2-6, 67-3-2 and 79-11-1 are placed in the middle of the neighbor-net presumably due to missing data for locus P1108-p. Cluster B consists of lineages from crosses 63, 67, 75 and 79 along with their *T. pratensis* parents. This cluster is further branched because of variation at locus P1116-p for particular S_0 lineages (79-1, 79-9 and 79-11) and missing data for their S_1 lineages (79-1-1,79-1-3, 79-11-1 and 79-11-3).

3.4.3.2 *T. mirus*

As with *T. miscellus*, three neighbor-nets also were constructed for the *T. mirus* synthetic polyploids. The first neighbor-net was based on all 16 progenitor-specific loci and included only the *T. mirus* synthetic lineages (Figure 3.3a). The second neighbor-net included the available *T. dubius* parents and the 'generic Types' genotypes only using the *T. dubius* loci (Figure 3.3b). These two neighbor-nets qualitatively agree but demonstrate greater distinction than the corresponding two *T. miscellus* neighbor-nets. As we here consider the synthetics relative to the Types I-III genotypes, the results covered in this section relate to the neighbor-net in Figure 3.3b. Again, all *T. mirus* synthetic lineages (F_1, S_0 and S_1 generations) produced from a single cross clustered together (with the odd exception of 99-2); this is

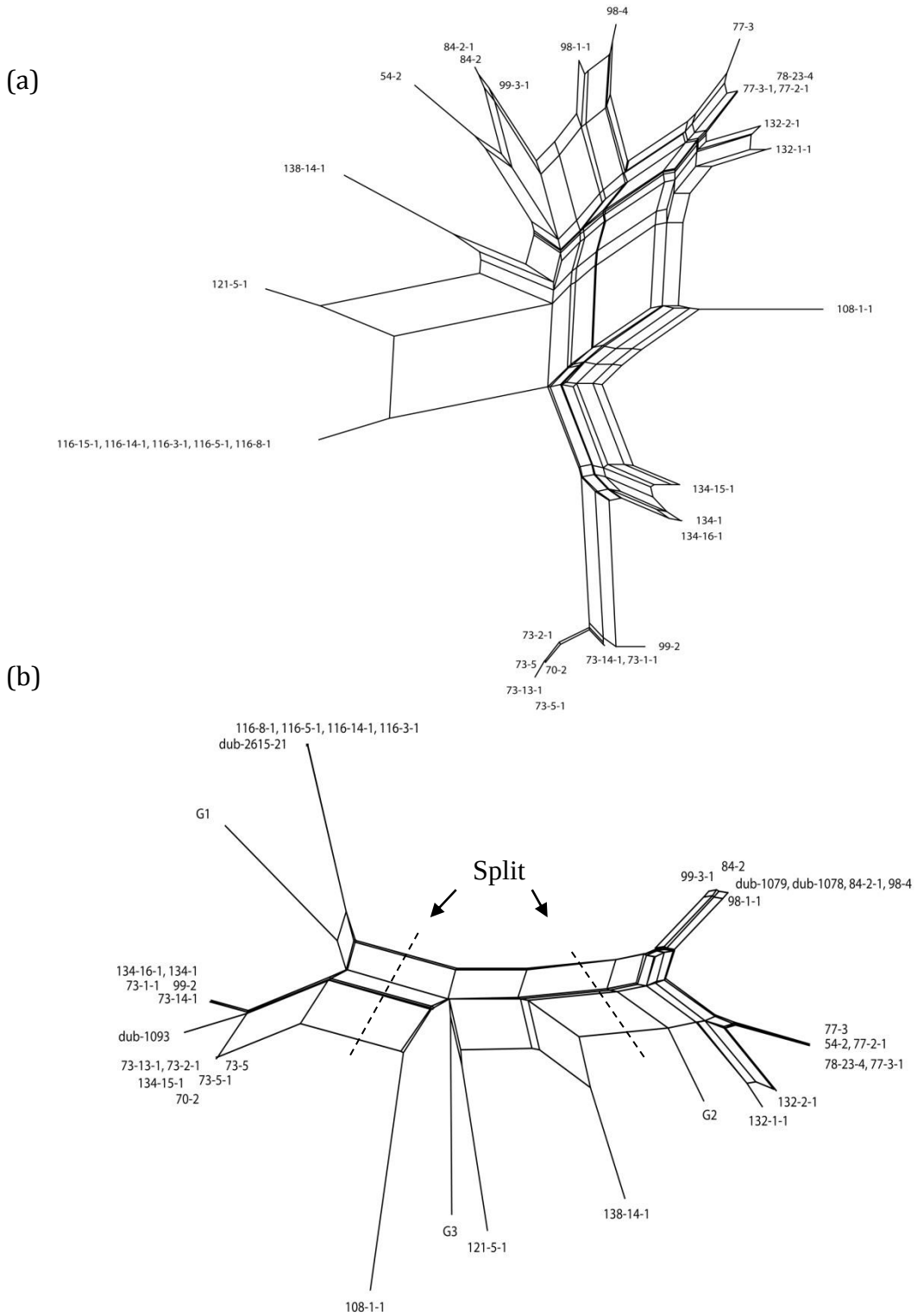


Fig. 3.3 Neighbor-net of *T. mirus* synthetic polyploids. (a) Based on all microsatellite loci (b) Based on only *T. dubius* loci and include their *T. dubius* parents and generic types named as G1, G2 and G3 based on multilocus genotype I, II and III. Each split corresponds to the cluster of individuals having distinct multilocus genotypes.

(a)

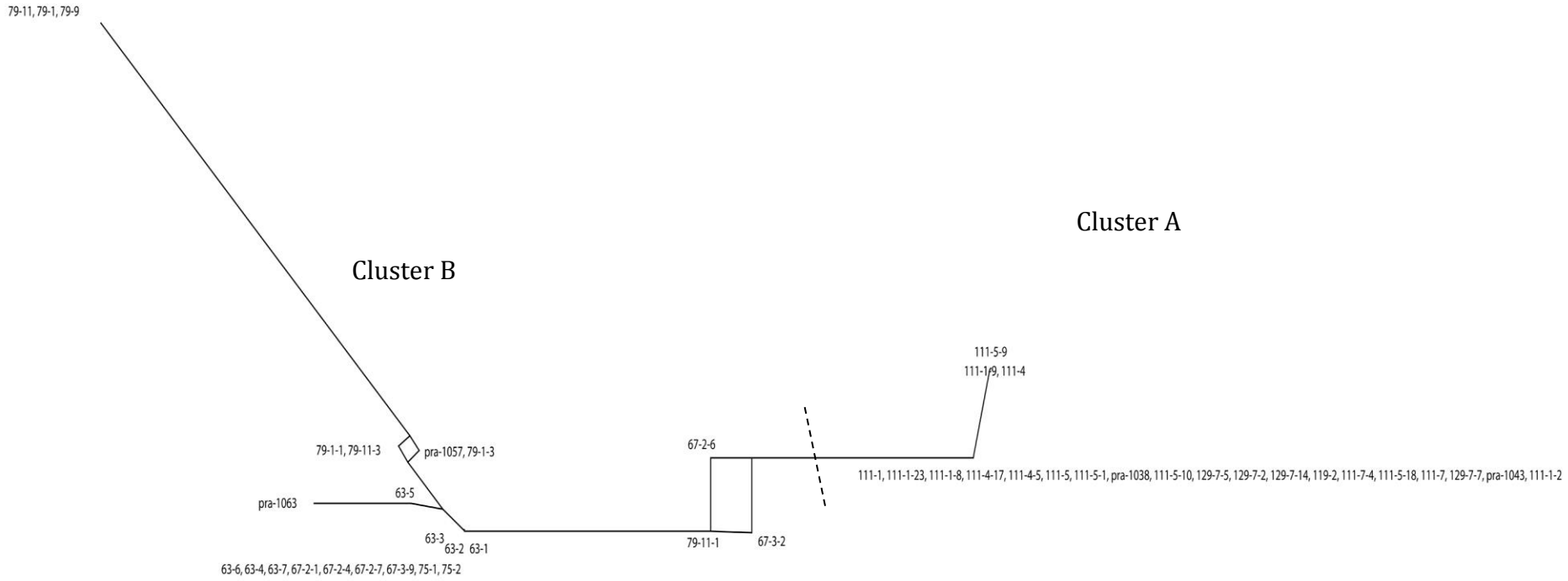


Fig. 3.4 (a) Neighbor-net of *T. miscellus* polyploids along with their *T. pratensis* parents.

(b)

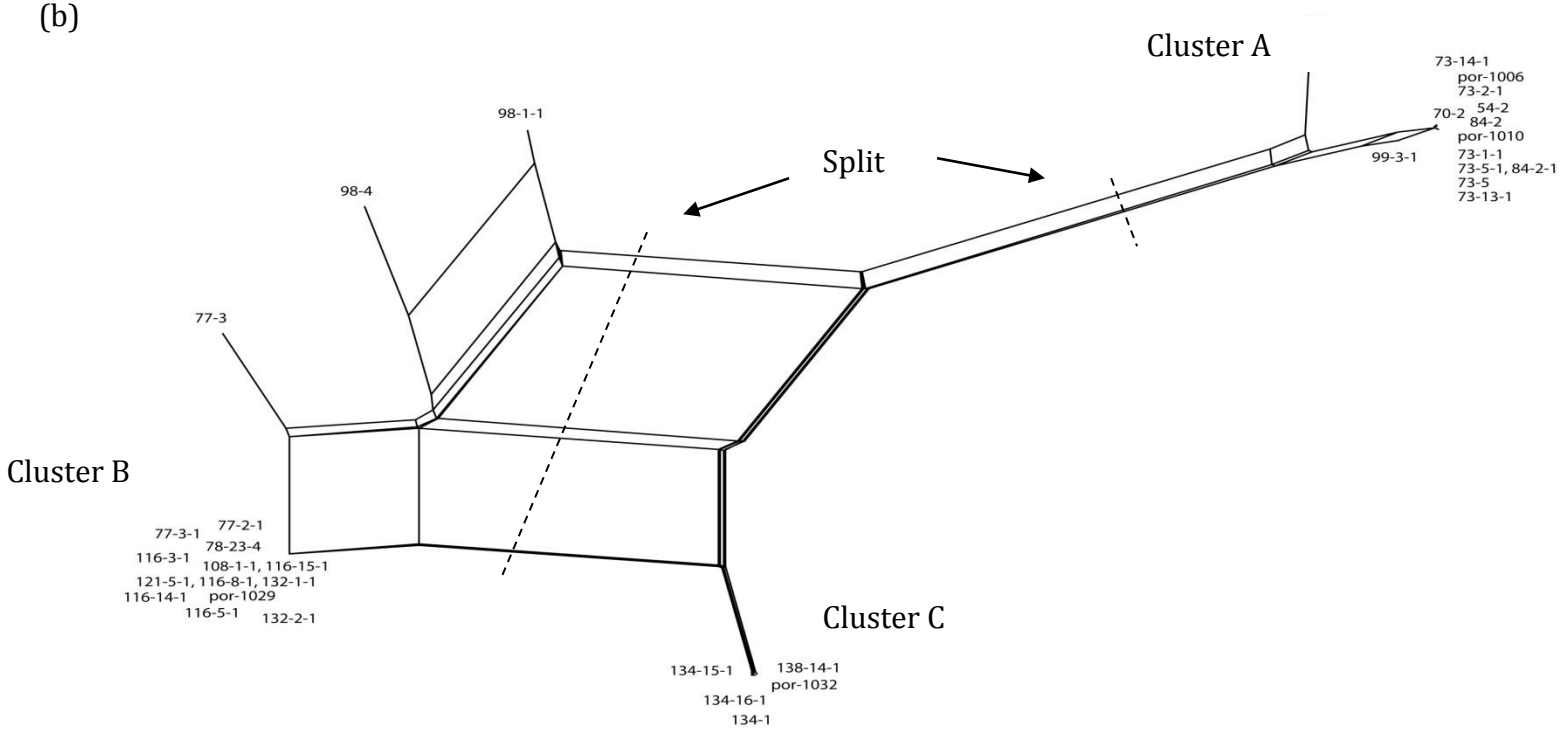


Fig. 3.4 (b) Neighbor-net of *T. mirus* polyploids along with their *T. porrifolius* parents.

because all of the *T. dubius* parents included in the analyses except 1093 (2613-24) were homozygous at all loci. One split separates lines 70, 73, 116 and 134 along with generic Type I from the rest of the network. These lines match the Type I five-locus genotype except at marker P1112-d; however none of these lines match the Type I genotype at more than five loci. Lineages from seven crosses are relatively good matches with the Type II genotype. Crosses 84, 98 and 99 match Type II at seven out of nine *T. dubius* loci, differing at loci D1054 and D1055. Lineages from crosses 54, 77 and 78 match this Type at all *T. dubius* loci except P1112-d, and line 132 matches the Type II at all *T. dubius* loci except D1048. Lineages from three crosses, 108, 121 and 138, had distinct genotypes from the other crosses and from the three Types. None of the synthetic *T. mirus* lineages included in the analyses were good matches for the Type III genotype, which is located central to the neighbor-net. The neighbor-net for synthetic *T. mirus* using all loci (Figure 3.3a) distinguishes nearly all synthetic lineages, except 77 and 78.

The final neighbor-net for synthetic *T. mirus* was constructed along with their *T. porrifolius* parents based on 6 loci (Figure 3.4b). This network could be divided into 3 clusters. One strongly supported split separates cluster 'A' containing lines 54, 70, 73, 84 and 99 along with their *T. porrifolius* parents 1006 (2611-3) and 1010 (2611-8) from cluster 'B' containing lines 77, 78, 108, 116, 121 and 132 along with *T. porrifolius* individual 1029 (2607-18) based on the variation at 5 out of 6 *T. porrifolius* loci. The *T. porrifolius* parents for each of these crosses were all from the same population. Cluster 'C', containing lines 134 and 138 along with their common *T. porrifolius* parent 1032 (2607-21), differ from its closest cluster 'B' at two *T. porrifolius* loci (P1112-p1 and P1112-p2). Lines 98-4 and 98-1-1 are distinct from all other lineages due of variation at loci P1110-p and P1112-p2.

3.5 Discussion

In the present study, we genetically characterize synthetic polyploids of *Tragopogon miscellus* and *T. mirus* using nuclear microsatellite markers designed from their diploid progenitors. To determine the genetic architecture of the synthetic polyploids and to confirm the genetic contribution of their diploid parents, several neighbor-nets were constructed. The results indicate Mendelian inheritance of all the parental loci in synthetic polyploids and the presence of multilocus genotypes (I-III) in the synthetics which were previously observed in natural polyploids (Symonds *et al.*, 2010). However, the location of those loci covered in these multilocus genotypes is not known in the *Tragopogon* genome.

3.5.1 Genetic structure of synthetic polyploids at microsatellite level

The genetic constitution of the synthetic *Tragopogon mirus* and *T. miscellus* polyploids was determined by using microsatellite markers. For both synthetic polyploids neighbor-nets were produced based on either all loci or loci from one parental species in combination with samples from that progenitor. Artificially produced generic genotypes (based on all the ten *T. dubius* loci), representing Type I, II and III found in the natural *Tragopogon* polyploids, were also included in the analysis to identify the frequency of those multilocus genotypes or similar ones in the synthetic polyploids. The structure observed in neighbor-nets that included *T. dubius* loci (alone or in combination with the alternative parent's loci) tended to reflect the three predominant Types previously observed in natural *T. miscellus* and *T. mirus*. Most of the synthetic lineages at least partially matched the multilocus genotypes (one marker D1112-d differed consistently). Of all the multilocus genotypes (based on the five diagnostic markers (see introduction)), Type I was present in both *T. mirus* and *T. miscellus* synthetic polyploids, Type II was present at higher frequency in *T. mirus* synthetics and Type III was present only in *T. miscellus* synthetic polyploids. The occurrence of those genotypes in synthetic polyploids is

plausible because their parents carried and incorporated those genotypes into the synthetic lines. Of all the multilocus genotypes, Type II was most supported as it matched the original multilocus genotype at 7 out of 10 *T. dubius* loci for most of the synthetic polyploids and their parents. Type III was absent in *T. mirus* synthetic lines because none of their *T. dubius* parents included in the analysis had that genotype.

Most of the synthetic polyploids matched the Types based on only five *T. dubius* loci and it might be unlikely to expect them to have multilocus genotypes that match the Types present in the natural polyploids at all ten *T. dubius* loci. The high frequency of the Types in natural *T. miscellus* and *T. m*irus may be due to ancestral population structure in *T. dubius* due to the introduction of different genotypes to the Palouse region (Symonds *et al.*, 2010). Following dispersal and gene flow among the different *T. dubius* introductions, that structure would have broken down such that modern *T. dubius* individuals are a mix of the original genotypes. Therefore, synthetic polyploids made by crossing modern *T. dubius* with *T. porrifolius* and *T. pratensis* would be unlikely to possess gentoypes that exactly match those of the natural polyploids, which formed long ago. Indeed, Symonds *et al.* (2010) found only one or two *T. dubius* individuals per multilocus genotype (I, II or III), while in the natural polyploids, nearly all individuals matched to those multilocus types at all 10 *T. dubius* loci. Moreover, very few *T. dubius* individuals have been used for producing synthetic polyploids by Tate *et al.* (2009b), so the presence of multilocus genotypes covering all 10 *T. dubius* loci might not be expected in the synthetic polyploids.

3.5.2 Genetic contribution of parental diploids into synthetic polyploid lineages

Several genetic changes take place post-polyploidization, which may cause imbalance of parental genomes in polyploids, and hence may affect the inheritance patterns of molecular markers (Lim *et al.*, 2008). Since *T. dubius* is a common parent of both synthetic polyploids, *T. mirus* and *T. miscellus*, one neighbor-net based on only *T. dubius* loci was constructed for each polyploid species in which *T. dubius* parents clustered

with their synthetic polyploid lineages (Figure 3.2b and 3.3b). As discussed earlier (see results), some of the *T. dubius* parents included in this study were heterozygous at some microsatellite loci and that heterozygosity was reflected in the clustering pattern of synthetic polyploids in the neighbor-nets. For instance, cross 111 has formed two small clusters because its *T. dubius* parent 1093 (2613-24) was heterozygous at some loci, whereas cross 63 has 1280 (2616-4) as its *T. dubius* parent, which was homozygous at all loci, so all the lineages of cross 63 were clustered together in the neighbor-net (Figure 3.2).

Although most synthetic lineages were at least partial matches for one of the three multilocus Types observed in natural *T. mirus* and *T. miscellus*, some lineages were quite different. Three *T. mirus* (cross 108, 121 and 138) and two from *T. miscellus* (cross 129 and 75) did not match with any multilocus genotype based on five *T. dubius* loci, hence did not cluster with either of those genotypes. Out of those that did not fit into those multilocus genotypes, only for cross 129 was the exact *T. dubius* parent, 1093 (2613-24), included in the study. These lineages were simply formed from *T. dubius* individuals that have putatively admixed genomes relative to the original *T. dubius* introductions.

Another illustration could be made from the clustering pattern of synthetic polyploids with their *T. dubius* parents, based on only *T. dubius* loci (Figure 3.2b and 3.3b). In the net, two clusters of the synthetic lineages were separated by split I and split II of the neighbor-net. Synthetic polyploid lineages of both of these clusters were produced by *T. dubius* parents from the Pullman population, except cross 119 and 116 whose *T. dubius* parents were from the Spokane population. Synthetic polyploid lineages of the third big cluster separated by split III were parented by *T. dubius* individuals from the Spangle population. As these splits I, II and III correspond to multilocus genotypes I, II and III respectively, based on these results it could be said that in synthetic polyploids multilocus genotype I and II were contributed by *T. dubius* individuals from the Pullman population while genotype III was contributed by *T. dubius* individuals from Spangle. Those synthetic polyploid crosses that did not match any multilocus genotype (e.g., cross 121 and 108) or

did not produce successful polyploid lineages (cross 75 and 119) were found to be produced by *T. dubius* individuals mostly from Spokane. These findings are plausible because out of 20 synthetic crosses of *T. mirus* and *T. miscellus* included in the present study, 11 crosses were parented by *T. dubius* individuals from Pullman, five crosses were parented by *T. dubius* individuals from Spokane and four crosses were parented by *T. dubius* individuals from Spangle. Hence, it could be assumed that *T. dubius* individuals from Pullman have contributed to making successful synthetic polyploids, and consequently contributing multilocus genotypes I and II which were most predominant in synthetic polyploids.

The contribution of the other two diploid parents to synthetic polyploids of *T. mirus* and *T. miscellus* could be observed from their neighbor-nets constructed separately based on only *T. porrifolius* or *T. pratensis* loci, respectively (Figure 3.4a and b). The results indicate that more variation is contributed by *T. porrifolius* than *T. pratensis* in synthetic polyploids, because *T. pratensis* was variable at only two of six loci analyzed in *T.miscellus*, while *T. porrifolius* was variable at all of the loci analyzed in *T. mirus* lineages. This observation is in parallel to the presence of more variation in *T. porrifolius* (Palouse populations) compared to *T. pratensis* populations in nature (Soltis *et al.*, 1995a).

3.5.3 Implication of genetic variation present in synthetic polyploids on the genetic changes observed in the synthetics

Based on the genetic characterization of synthetic polyploids of *Tragopogon* and the analysis of the genetic variation contributed by the diploid parents, inferences can be drawn for the genetic changes observed by previous research in these *Tragopogon* synthetic polyploids. Our results have shown Mendelian inheritance of all microsatellite loci, similar to Buggs *et al.* (2009), who found complete additivity with no evidence of homeolog loss in 44 S_1 individuals of *T. miscellus* synthetic polyploids.

Most previous research on *Tragopogon* synthetic polyploids has focused on karyotype studies and evolution of rDNA loci. Lim *et al*. (2008) found meiotic irregularities

(multivalent formation) in crosses 73 and 134 of *T. mirus* synthetic polyploids and variable pattern of rDNA loci. Tate *et al*. (2009b) found meiotic abnormalities in lines 70-4, 98-1 and 132-1, while other polyploids (line 67-3) showed both normal bivalent formation as well as some chromosome pairing abnormalities. Malinska *et al.* (2010) observed copy number variation of 18S rDNA loci and deviation from additive parental ratios in some *T. mirus* crosses (70, 73 98, 134 and 135), which were skewed towards *T. porrifolius*. According to our results, these lineages belong to either multilocus genotype I or II. Noticeably, most of the meiotic abnormalities and biases in expression pattern of rDNA loci were observed in only *T. mirus* synthetic polyploids and interestingly the source of parental genomes for most of *T. mirus* synthetics was the Pullman population (Tate *et al.*, 2009b). There could be some specific genotypes that more successfully formed polyploids in the wild and this may be the case with the synthetic polyploids. It seems that *T. dubius* individuals from Pullman were more successful in generating *T. mirus* synthetic polyploids.

Here we characterized the genetic variation present within and among lineages of synthetic *T. miscellus* and *T. mirus*. Our results indicate that considerable variation has been contributed to the collective synthetics although a more thorough analysis that includes simultaneous characterization of the synthetics along with the natural *Tragopogon* polyploids is warranted. The data indicate that the multilocus Types identified in the natural polyploids are somewhat preserved in the synthetics, but only as much as is possible given the admixture in natural *T. dubius* due to its multiple introductions. Interestingly, the synthetic genotypes that most deviate from the Types that occur in natural *T. miscellus* and *T. mirus* appear to have had the least success in advancing beyond the F_1 and S_0 generations. Unfortunately, without detailed information on how/why these lineages did not advance, little can be inferred; however, it is an intriguing finding that may warrant a more detailed and controlled investigation that would likely require the production of more synthetics. Finally, the simple characterization of the relationships of the existing *T. miscellus* and *T. mirus* synthetic lineages makes a strong contribution to the

comparative power available to those investigating patterns of genomic reorganization and patterning using these lines.

CHAPTER FOUR

4. Comparative analysis of floral transcriptomes

4.1 Abstract

Tragopogon miscellus has formed reciprocally exhibiting short-liguled (*T. dubius*♂ × *T. pratensis*♀) and long-liguled (*T. dubius*♀ × *T. pratensis*♂) forms. As these morphologies were repeated in synthetic polyploids, here we propose a maternal influence to the formation and variation in floral morphology. To investigate the genetic factors controlling these floral morphologies, floral transcriptomes were studied from diploid parents and reciprocally formed natural and synthetic *T. miscellus* polyploids. Orthologs of MADS-box transcription factors and other floral developmental genes were identified in *Tragopogon* for the first time. Considerable variation has also been detected within both parental species for the floral identity genes. Unfortunately, polymorphisms between parents were not informative for determining progenitor-specific expression in the polyploids. As this was a preliminary study, a comprehensive analysis of this transcriptomic data would yield valuable findings. Moreover, a thorough developmental study of the individual floral whorls and inflorescences is required to identify the genes responsible for these differing inflorescence forms.

4.2 Introduction

Angiosperms represent the largest and morphologically most diverse group of plants (APGIII, 2003; Soltis & Soltis, 2004). The basic floral structure consists of four whorls: the first outer whorl consists of sepals, the second whorl contains petals, the third whorl contains stamens (androecium), and the fourth central whorl has carpel(s) (gynoecium) (Meyerowitz *et al.*, 1991). Flowering takes place as a sequence of events involving a switch from a vegetative to a reproductive phase by floral meristem identity genes and then the formation of floral organ primordia by the flower development genes (Krizek & Fletcher, 2005). The

transition from a vegetative to a reproductive phase is important for flowering to occur at an appropriate time in order to synchronize with climatic conditions, such as wind-pollination or the availability of pollinators for cross-pollinated species (Huijser & Schmid, 2011). Floral initiation begins by the activation of inflorescence meristem identity genes at the shoot apical meristem (Krizek & Fletcher, 2005). In *Arabidopsis* this induction is brought about by *LEAFY* and *APETALA1,* while this function in *Antirrhinum* is accomplished by *FLORICAULA* (*FLO*) and *SQUAMOSA* (*SQUA*) (Coen *et al.*, 1990; Huijser *et al.*, 1992; Weigel *et al.*, 1992; Mandel & Yanofsky, 1995). Floral morphogenesis in angiosperms is controlled by a combination of A, B, and C class genes of the ABC model proposed by Coen and Meyerowitz (1991). These gene classes determine the fate of developing floral organ primordia (Ng & Yanofsky, 2000). Floral identity genes were first characterized in *Arabidopsis thaliana* via molecular cloning and mutant analysis (Komaki *et al.*, 1988; Bowman *et al.*, 1989; Bowman *et al.*, 1991) and their orthologs have been widely reported in *Petunia* and *Antirrhinum majus,* as well as many other eudicots and some monocots, e.g., rice and maize (Stubbe, 1966; Sommer *et al.*, 1990; Angenent *et al.*, 1992; Huijser *et al.*, 1992; Saedler & Huijser, 1993; Tsuchimoto *et al.*, 1993; Vanderkrol & Chua, 1993; Kang *et al.*, 1998; Munster *et al.*, 2001; Albert *et al.*, 2005; Whipple *et al.*, 2006; Soltis *et al.*, 2007). These floral identity genes are part of the MADS family, where MADS (MCM1, Agamous, Deficiens, Serum response factor) is a conserved DNA-binding domain present in transcription factors and represents one of the largest multi-gene families in plants (Riechmann & Meyerowitz, 1997).

ABC gene classes are considered to be homeotic as they determine the sequence of events for the developmental pattern of floral organs. Any mutation in them causes changes in the pattern of floral parts (Coen & Meyerowitz, 1991). A-class homeotic genes control sepal development in the first whorl (Fig. 4.1) and in *Arabidopsis thaliana* these genes are *APETALA1* (*AP1*) and *APETALA2* (*AP2*), while in *Antirrhinum majus* an *AP2*-like *LIP1* and *LIP2*

perform class A function (Mandel *et al.*, 1992; Jofuku *et al.*, 1994; Keck *et al.*, 2003). Combined with A-class function, class B genes control petal formation in the second whorl (Fig. 1). In *A. thaliana*, these are *APETALA3* (*AP3*) and *PISTILLATA* (*PI*), while their orthologs in *A. majus* are *DEFICIENS* (*DEF*) and *GLOBOSA* (*GLO*) (Sommer *et al.*, 1990; Trobner *et al.*, 1992; Goto & Meyerowitz, 1994; Jack *et al.*, 1994). Combined with B-class genes, class C genes control stamen formation in the third whorl (Fig. 4.1). C-class genes are antagonistic and have a repressive role to A-class gene function. C-class in *A. thaliana* includes *AGAMOUS* (*AG*) and in *A. majus PLENA* and *FARINELLI* (*FAR*) specify class C function (Yanofsky *et al.*, 1990; Bradley *et al.*, 1993; Davies *et al.*, 1999). C-class genes alone control carpel identity in the fourth whorl (Fig. 1). Additions to the ABC model include D-class genes, which control ovule identity inside carpels. These were first identified in *Petunia hybrida; FLORAL BINDING PROTEIN* (*FBP7* and *FBP11*) controls ovule formation in *Petunia* (Cheng *et al.*, 2000), while in *Arabidopsis* ovule development is controlled by *SHATTERPROOF 1* and *2* and *SEEDSTICK* (*STK*) genes (Pinyopich *et al.*, 2003). E-class genes (*SEPALLATA*) are required at all stages in *Arabidopsis*, but their requirement in other groups is not yet known. In *Petunia*, class E function is performed by *FBP2* and *FBP7* genes (Pelaz *et al.*, 2000; Ferrario *et al.*, 2003; Vandenbussche *et al.*, 2003).

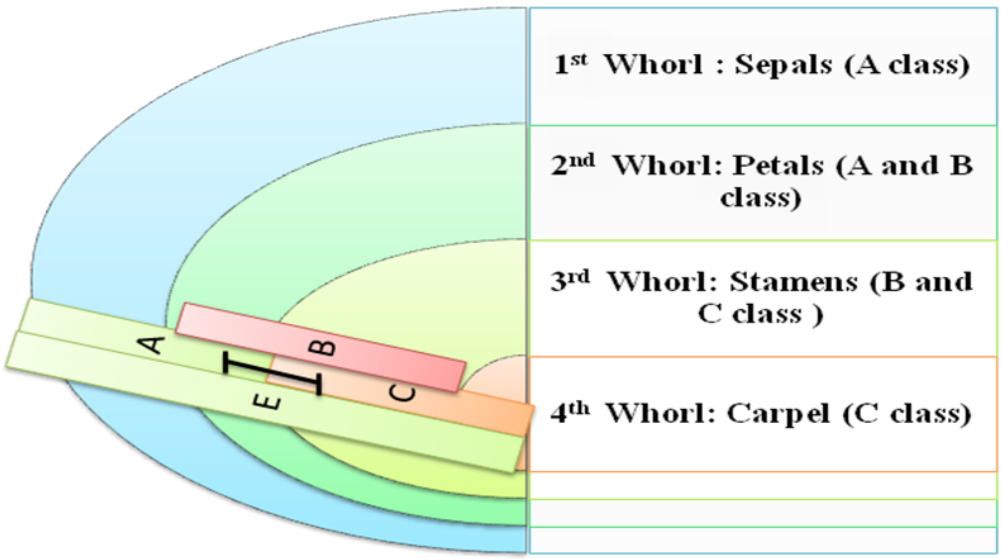


Fig. 4.1 Diagrammatic illustration of the ABC model.

In angiosperms, floral symmetry (zygomorphy and actinomorphy) is brought about by a group of *CYCLODIA* (*CYC*-like) genes which belong to the class II TCP transcription factor family (Hileman & Cubas, 2009). In *Antirrhinum majus*, which is bilaterally symmetrical (zygomorphic), *CYC*-like genes and their paralogs *DICHOTOMA* are expressed dorsally in the flower controlling petal size, shape and stamen primordium. Mutations in *CYC*-like genes result in radial symmetry in *Antirrhinum (Cubas, 2004)*. On the other hand, in *Arabidopsis*, a *CYC*-like gene is also dorsally expressed, resulting in radially symmetrical flowers. This suggests that changes in spatial expression and interaction with other genes may be required to form bilateral symmetry by *CYCLODIA*-like genes (Coen *et al.*, 1995; Cubas, 2004).

Inflorescence structure in the family Asteraceae is comprised of composite flowers where florets are arranged in a capitulum or head. In many species, each head contains ligulate ray florets (zygomorphic) arranged in the outer whorl and tubular disc florets (actinomorphic) present in the centre of the head, while some species have only ray or disc florets. Sex expression in different floret types is variable. Disc florets are usually hermaphroditic (both male and female fertile), while ray florets are usually female fertile and have staminodes (non-functional stamens). The entire inflorescence is enclosed in one or more whorls of green receptacular bracts (Koch, 1930; Bremer, 1994; Soltis *et al.*, 2006; Patterson, 2009). The existence of morphologically and functionally diverse floret types in a single inflorescence makes the Asteraceae an interesting group to study expression regulation of the genes involved in flower morphogenesis. Floral identity genes have been explored in a few members of Asteraceae including *Helianthus annuus*, *Chrysanthemum* and *Gerbera hybrida*. Dezar *et al*., (2003) first identified MADS-box genes in sunflower. Later, Shulga *et al*. (2008) reported the genetic basis of floral morphogenesis in sunflower. They isolated and characterized the orthologs of the ABC-classes genes previously reported in *Arabidopsis* and *Antirrhinum*. In *Chrysanthemum*, Shchennikova *et al*. (2004) studied floral development genes and elucidated

their functional characterization based on protein-protein interactions and complementation analysis. They found that orthologs of floral identity genes in *Chrysanthemum* were functionally equivalent to *AP1* and *SEP3* genes from *Arabidopsis*. In *Gerbera hybrida,* Yu *et al*. (1999) described the expression pattern of floral identity genes, their function and phylogenetic relationship with previously reported MADS-box genes. They also showed that the pattern of induction of floral identity genes in *Gerbera* was spatio-temporally typical of the head-like inflorescence found in Asteraceae. Based on studies of transgenic plants produced for MADS box genes isolated from *Gerbera*, they reported that pappus bristles in *Gerbera* are modified sepals responsible for seed dispersal.

Although flower morphogenesis in Asteraceae has been widely studied at the morphological and developmental level (Bremer, 1994; Harris, 1995; Soltis *et al.*, 2006), limited work has been done at the molecular level (Yu *et al.*, 1999; Dezar *et al.*, 2003; Shchennikova *et al.*, 2004; Uimari *et al.*, 2004; Shul'ga *et al.*, 2008). Additional studies are needed in other genera to understand the molecular genetic basis of floral and inflorescence development. *Tragopogon* also belongs to family Asteraceae and has 150 species found in Eurasia, mostly comprising diploid species (2*n* = 12) and some polyploid species (2*n*= 24) (Ownbey, 1950). Three diploid species (*Tragopogon dubius*, *T. pratensis* and *T. porrifolius*) have been commonly reported in northwestern America. These diploid species have repeatedly formed two polyploid species: *T. miscellus* has formed reciprocally (short-liguled form: *T. dubius*♂ × *T. pratensis*♀; long-liguled form: *T. dubius*♀ × *T. pratensis*♂) and *T. mirus* has formed in one direction naturally (*T. dubius*♂ × *T. porrifolius*♀) (Ownbey & Mccollum, 1953; Soltis & Soltis, 1989; Soltis *et al.*, 2004). Recently synthetic polyploids of both *T. miscellus* and *T. mirus* have also been produced by Tate *et al*. (2009b). These synthetic polyploids morphologically resemble the natural polyploids in terms of inflorescence shape, ligule size and colour (Tate *et al.*, 2009b) and represents an ideal system for a comparative

study of floral and inflorescence development. Given that differences in floral morphology are observed repeatedly in the reciprocally formed *T. miscellus* polyploids, some maternal influence on the floral development is suggested.

Here, I conducted a pilot study of the expression of MADS box floral transcription factors and their potential role in determining inflorescence and floral morphology in *Tragopogon*. For this, I sequenced floral transcriptomes of reciprocally formed natural and synthetic *T. miscellus* polyploids along with their diploid parents (*T. dubius* and *T. pratensis*) to answer: (1) How diverged are *T. dubius* and *T. pratensis* for floral developmental genes and what is the level of expression of progenitor-specific transcripts in reciprocally formed *T. miscellus*? (2) Is there any variation between short- and long-liguled *T. miscellus* at the transcriptome level that could correspond to their floral morphologies?

4.3 Materials and methods

4.3.1 Plant Material

Seeds of the two diploid parental species, *Tragopogon dubius* and *T. pratensis,* as well as short- and long-liguled natural *Tragopogon miscellus*, were previously collected from natural populations in Pullman, Washington and Moscow, Idaho, USA and were grown in the glasshouse at Massey University under standard conditions. Synthetic first-generation polyploids of *T. miscellus* produced previously by Tate *et al*. (2009b) were also grown from seed in the glasshouse (Table 4.1). Both naturally formed and synthetically produced reciprocal *T. miscellus* polyploids with long (*T. dubius*♀ × *T. pratensis*♂) and short ligules (*T. dubius*♂ × *T. pratensis*♀) were included in the study (Table 4.1).

Table 4.1 List of diploid and polyploid samples.

Sample	Floral structure	Cross number	Unique ID	lineage
T. pratensis	 Diploid parent		1039	2608-1
T. dubius	 Diploid parent		1099	2613-50
T. miscellus	 Allotetraploid Natural-long liguled	490	1202	2605-9
T. miscellus	 Allotetraploid Natural-Short liguled	370	1766	2604-22
T. miscellus	 Allotetraploid Synthetic-long liguled	129-7	1093	
T. miscellus	 Allotetraploid Synthetic-short liguled	111-4	1038	

4.3.2 RNA extraction

For each species, RNA was extracted in two replicates containing young and mature floral buds. Both young (closed buds with a developing greenish capitulum) and mature (closed buds with mature ray and disk florets on the developed capitulum) floral buds were separately collected directly into liquid nitrogen after removing bracts. RNA was extracted from 100 mg of floral material using Qiagen RNeasy Mini kit (Hilden, Germany). The first step of the extraction protocol was modified by adding 4% polyvinyl pyrrolidone (PVP) in the lysis buffer to remove phenolic compounds. After RNA extraction, RNase inhibition and DNAse digestion were performed; RNA (5-50 µg) was incubated with 20 units of Protector RNase inhibitor (Roche, Germany), 10 units of DNase I and 10X DNase I incubation buffer in a total volume of 50 µl at 37 °C for 15 min. The reaction was stopped by adding EDTA to a final concentration of 8mM and incubated at 75 °C for 10 min. RNA was ethanol-precipitated in the presence of 3M sodium acetate and re-dissolved in 25 µl of RNase-free water.

4.3.3 RNA quantification and quality control

RNA was quantified using a Nanodrop-1000 spectrophotometer (Thermo Fisher Scientific, USA). Optical density ratios for 260/280 and 260/230 = 2 or more were preferred for evidence of good quality RNA samples that were free of DNA contamination, respectively. Integrity of RNA was tested by running total RNA samples on an Agilent 2100 Bioanalyzer using the RNA-6000 Nano LabChip kit at Massey Genome Service (Palmerston North, New Zealand). RNA samples with RNA Integrity Number (RIN) of more than 8 and rRNA ratio (28S/18S) of 1.5-2.5 were submitted for RNA sequencing (Illumina Miseq and Hiseq) at New Zealand Genomics Limited (NZGL).

4.3.4 RNA Sequencing

Upon submission to the NZGL, RNA samples were quantified using a Qubit Flourometer (Invitrogen, USA). Quant-iT RNA Assay was performed by Massey Genome Service (Palmerston North, New Zealand) to determine if samples had the required concentration of more than 200ng/µl of total RNA for each sample. Six sample libraries (*T. dubius, T. pratensis*, natural short- and long-liguled *T. miscellus* and synthetic short- and long-liguled *T. miscellus*) were prepared using Illumina TruSeq RNA library preparation kit (Illumina, USA). Integrity of the libraries was checked on an Agilent 2100 Bioanalyzer to determine adapter or primer dimer contamination and for the presence of secondary products. Libraries were co-loaded onto a single flow cell and run on an Illumina Miseq for a 2X 250 bp paired-end run at Massey Genome Service (Palmerston North, New Zealand). These same libraries were then co-loaded on a single flow cell on the Illumina Hiseq (Otago University, Dunedin, New Zealand) for a 2X 100 bp paired-end run. This strategy was employed to combine longer-read transcripts (MiSeq) with greater sequencing depth (HiSeq) for *Tragopogon*, which does not have a reference genome available.

4.3.5 Data analysis

MiSeq and HiSeq data were quality checked by New Zealand Genomics Limited (NZGL). For this purpose, FastqMcf Version 1.04.636 was used from the package Ea-util for adaptor trimming (Aronesty, 2013), FastqScreen was used to check for contaminants (genomes of other species being used in a laboratory, PhiX (adapter-ligated library that is used as a control in Illumina experiment), and vectors). Additional quality checks were kindly performed by Murray Cox using SolexaQA package (Cox *et al.*, 2010) to quality trim (P = 0.05) the reads. After cleaning, reads fewer than 40 base pairs were discarded and the remaining reads were mapped to the set of reference genes. As we did not have a *Tragopogon* reference genome for mapping purposes, the reference genes used here include previously characterized floral

development genes from close relatives of *Tragopogon* from the Asteraceae (Table 4.2). After mapping RNA sequence HiSeq and MiSeq reads to the reference genes, their alignments were viewed in Integrative Genome Viewer (IGV) version 2.3 (Robinson *et al.*, 2011; Thorvaldsdóttir *et al.*, 2013). Based on the polymorphisms present between the diploid parental species (*T. dubius* and *T. pratensis*), for each reference gene the number of progenitor-specific transcripts mapped were counted for the natural and synthetic *T. miscellus* reciprocal polyploids.

4.4 Results

In order to determine the level of expression of floral development genes, transcriptome reads from both MiSeq and HiSeq runs for diploid parents (*Tragopogon dubius* and *Tragopogon pratensis*) and their natural and synthetic polyploid species *Tragopogon miscellus* were mapped to orthologous floral development genes from other Asteraceae species (*Helianthus annuus*, *Chrysanthemum*, *Gerbera hybrida* and *Lactuca sativa*). On an average, MiSeq run produced more than 3 million reads per sample and HiSeq run produced more than 20 million reads per sample.

4.4.1 Divergence between parental species

The overall level of heterozygosity was similar in *Tragopogon dubius* and *T. pratensis* for the orthologs of the floral identity genes; although *T. pratensis* had slightly more SNPs than *T. dubius*. Divergence between the parental species was determined from progenitor-specific polymorphisms in the floral development genes. These polymorphisms were used to determine the level of expression of parental homeologs in *T. miscellus* polyploids. At single nucleotide polymorphisms (SNPs) where the two parents differed, *Tragopogon dubius* was usually homozygous, while *T. pratensis* was typically heterozygous at those SNPs.

Table 4. 2 List of MADS-box orthologs from other species of Asteraceae that were used as reference genes.

Species	Gene	Accession number	Reference
Sunflower	*HAM137*	AAO18233	Shulga *et al*., 2008
Sunflower	*HAM2*	EF612597	Shulga *et al*., 2008
Sunflower	*HAM31*	AAO18230	Shulga *et al*., 2008
Sunflower	*HAM45*	AAO18228	Shulga *et al*., 2008
Sunflower	*HAM59*	AAO18229	Shulga *et al*., 2008
Sunflower	*HAM63*	EF612598	Shulga *et al*., 2008
Sunflower	*HAM75*	AAL83209	Shulga *et al*., 2008
Sunflower	*HAM91*	AAO18231	Shulga *et al*., 2008
Sunflower	*HAM92*	AAO18232	Shulga *et al*., 2008
Sunflower	*HaAP3*	AY185363.1	Dezar *et al*., 2003
Sunflower	*HaAG*	AY157724	Dezar *et al*., 2003
Sunflower	*HaPI*	AY157725	Dezar *et al*., 2003
Sunflower	*Cycloidea*-like 1a	EU088366	Chapman *et al*., 2008
Sunflower	*Cycloidea*-like 1b	EU088367	Chapman *et al*., 2008
Sunflower	*Cycloidea*-like 2a	EU088368	Chapman *et al*., 2008
Sunflower	*Cycloidea*-like 2b	EU088369	Chapman *et al*., 2008
Sunflower	*Cycloidea*-like 2c	EU088370	Chapman *et al*., 2008
Sunflower	*Cycloidea*-like 2d	EU088371	Chapman *et al*., 2008
Sunflower	*Cycloidea*-like 2e	EU088372	Chapman *et al*., 2008
Sunflower	*Cycloidea*-like 3a	EU088373	Chapman *et al*., 2008
Sunflower	*Cycloidea*-like 3b	EU088374	Chapman *et al*., 2008
Sunflower	*Cycloidea*-like 3c	EU088375	Chapman *et al*., 2008
Sunflower	*FLOWERING LOCUS-T2*	GQ884982	Blackman *et al*., 2010
Chrysanthemum	*CDM111*	AY173054	Shchennikova *et al*., 2004
Chrysanthemum	*CDM115*	AY173060	Shchennikova *et al*., 2004
Chrysanthemum	*CDM19*	AY173064	Shchennikova *et al*., 2004
Chrysanthemum	*CDM36*	AY173065	Shchennikova *et al*., 2004
Chrysanthemum	*CDM37*	AY173059	Shchennikova *et al*., 2004
Chrysanthemum	*CDM41*	AY173055	Shchennikova *et al*., 2004
Chrysanthemum	*CDM44*	AY173057	Shchennikova *et al*., 2004
Chrysanthemum	*CDM8*	AY173056	Shchennikova *et al*., 2004
Chrysanthemum	*CDM86*	AY173061	Shchennikova *et al*., 2004
Chrysanthemum	Flowering time locus	JF488071	Tian *et al*., 2011
Gerbera hybrida	*Cycloidea* like 10	JN190064	Broholm *et al*., 2008
Gerbera hybrida	*Cycloidea* like 5	JN190059	Broholm *et al*., 2008
Gerbera hybrida	*Cycloidea* like 6	JN190060	Broholm *et al*., 2008
Gerbera hybrida	*Cycloidea* like 7	JN190061	Broholm *et al*., 2008
Gerbera hybrida	*Cycloidea* like 8	JN190062	Broholm *et al*., 2008
Gerbera hybrida	*Cycloidea* like 9	JN190063	Broholm *et al*., 2008
Gerbera hybrida	*GRCD1*	AJ400623	Kotilainen *et al*., 2000
Gerbera hybrida	*GRCD2*	AJ784156	Uimari *et al*., 2004
Gerbera hybrida	*gaga1*	AJ009722	Yu *et al*., 1999
Gerbera hybrida	*gaga2*	AJ009723	Yu *et al*., 1999
Gerbera hybrida	*gdef1*	AJ009724	Yu *et al*., 1999
Gerbera hybrida	*gdef2*	AJ009725	Yu *et al*., 1999
Gerbera hybrida	*gglo1*	AJ009726	Yu *et al*., 1999
Gerbera hybrida	*gsqua1*	AJ009727	Yu *et al*., 1999
Gerbera hybrida	*ghCyc4*	EU429305	Broholm *et al*., 2008
Gerbera hybrida	*ghCyc1*	EU429302	Broholm *et al*., 2008
Gerbera hybrida	*ghCyc2*	EU429303	Broholm *et al*., 2008
Gerbera hybrida	*ghCyc3*	EU429304	Broholm *et al*., 2008
Lactuca sativa	Flowering time locus	AB602323	Fukuda *et al*., 2011
Lactuca sativa	*Pistillata* like-Class B	Lsa012402	Zhang *et al*., 2011

For instance, at a G/A SNP, where all transcripts from *T. dubius* were G, then at that SNP, *T. pratensis* reads were A or G in near-equal proportions. In this example, only 'A' could be considered *T. pratensis*-specific because it was absent in *T. dubius*. However, these SNPs were not informative for determining progenitor-specific expression in the polyploids. Unfortunately, the above situation for SNPs between *T. dubius* and *T. pratensis* existed in >90% of the floral identity genes studied here, making inference about homeolog-specific expression difficult in *T. miscellus* polyploids.

4.4.2 Expression of floral development genes in *Tragopogon* diploids and polyploids

The total number of transcripts from diploid parents and *T. miscellus* polyploids mapping to floral identity genes from other Asteraceae species (Table 4.3) and comparative transcript abundance of floral identity genes in long- and short-liguled natural and synthetic polyploids were calculated (Table 4.4). Comparative transcript abundance was determined by calculating the percentage of transcripts mapping to orthologs from other species (Table 4.4). The minimum level of significant expression of either parental homeolog was set as ≥70%, meaning that a parental homeolog was considered to be significantly expressed only if its percentage was equal to or more than 70% at each diagnostic SNP.

4.4.2.1 Expression of A-class genes

A-class genes confer sepal identity and in sunflower, *HAM75* and *HAM92* perform A function (Shul'ga *et al.*, 2008). *Tragopogon* transcripts (from both MiSeq and HiSeq) orthologous to *HAM75* were polymorphic with three SNPs identified between the parental homeologs at 294 bp, 319 bp and 328 bp, respectively (these are relative positions on the orthologous gene). Natural short- and long-liguled (2604-22 and 2605-9, respectively) *T. miscellus* polyploids and also both short- and long-liguled synthetic polyploids (111-4 and

129-7, respectively) were heterozygous at all the SNPs (Table 4.4). For the second A-class gene from sunflower *HAM92*, there were again three SNPs identified at 374 bp, 382 bp, and 427 bp in *Tragopogon* orthologs. All the natural and synthetic *T. miscellus* polyploids equally expressed those polymorphic transcripts. In *Chrysanthemum CDM111*, *CDM8* and *CDM41* were reported to perform A-function (Shchennikova *et al.*, 2004). Three SNPs were identified in *T. miscellus* polyploids orthologs for both *CDM111* (433 bp, 480 bp and 580 bp) and *CDM41* (167 bp, 184 bp and 325 bp). None of the reads from parents or polyploids mapped to *CDM8*. For both *CDM111* and *CDM41*, all the reciprocally formed natural and synthetic polyploids showed almost an equal proportion of heterozygous transcripts. In *Gerbera hybrid*, *SQUAMOSA*-like *gsqua1* has been reported for A function (Yu *et al.*, 1999). Transcript reads orthologous to *gsqua1* abundantly mapped from both *Tragopogon* diploid parents and *T. miscellus* polyploids (Table 4.3). These *Tragopogon* orthologs were highly homozygous and conserved with not even a single SNP present in them.

4.4.2.2 Expression of B-class genes

In combination with A activity genes, B-class genes control petal formation in the second whorl and along with C-class genes, B-class genes are involved in the development of stamens in the 3rd whorl of the flower (Krizek & Fletcher, 2005). Expression patterns of the orthologs of sunflower B-function genes: *HAM2*, *HAM31*, *HAM63*, *HAM91*, *HaPI* and *HaAP3* (Shul'ga *et al.*, 2008) were determined in *Tragopogon* diploids and polyploids. *Tragopogon* transcripts orthologous to *HAM2* have three SNPs at 256 bp, 274 bp and 343 bp; orthologs of *HAM31* had one SNP at 496 bp; and orthologs of *HaPI* had one SNP at 153 bp. All the natural and synthetic *T. miscellus* polyploids equally expressed those polymorphic transcripts (Table 4.4). Very few *Tragopogon* transcripts mapped to *HAM63* and *HaAP3* and there was no polymorphism observed in them (Table 4.3). Orthologs of *HAM91* had five SNPs at 107 bp, 114 bp, 154 bp, 157 bp and 160 bp. The third SNP (154 bp) was unique in that all the natural

and synthetic polyploids were homozygous like *T. pratensis* (100% T) while *T. dubius* was heterozygous (G-92% and T-8%) at that SNP (Table 4.4). The remaining four SNPs seemed to be linked in having almost equal percentages of polymorphic transcripts at each of them in all *T. miscellus* polyploids except natural long-liguled *T. miscellus* (2605-9) which expressed *T. pratensis*-specific transcripts at a lower level (27-33%) compared to the other allele common between the parental species (67-73%) (Table 4.4). In *Chrysanthemum*, *PI*-like *CDM86* and *AP3*-like *CDM19* and *CDM115* perform B-function (Shchennikova *et al.*, 2004). *Tragopogon* transcripts orthologous to *CDM86* had one SNP at 500 bp and orthologs of *CDM19* had four SNPs at 193 bp, 461 bp, 529 bp and 570 bp. For both of these, *Tragopogon* orthologs equally expressed both polymorphic transcripts in all the *T. miscellus* polyploids studied except natural long-liguled *T. miscellus* (2605-9) which was homozygous (G-100%) for *CDM19* orthologs at 4th SNP (570 bp). Very few transcripts mapped to *CDM115* and there were no SNPs observed in the between *T. dubius* and *T. pratensis* orthologs of *CDM115*. In *Gerbera hybrida*, *GLO*-like *gglo1* and *DEF*-like *gdef1* and *gdef2* have been reported for B activity (Yu *et al.*, 1999). *Tragopogon* orthologs of *gglo1* had one SNP at 154 bp and were heterozygous with equal percentage of variants (50:50 both C and T variants) in all the polyploids and orthologs of *gdef1* had six SNPs at 77 bp, 84 bp, 124 bp, 127 bp, 129 bp, and 504 bp. All the natural and synthetic *T. miscellus* polyploids were heterozygous (Table 4.4). Similar expression patterns were observed for the orthologs of *gdef2* at the three SNPs at 163 bp, 185 bp, and 275 bp in all the *T. miscellus* individuals except one synthetic short-liguled individual (111-4) which showed higher expression of *T. pratensis* homeolog (81%) compared to other homeolog common between the parents (Table 4.4). *Tragopogon* orthologs of *Lactuca sativa PISTILLATA* like- B function gene had one SNP at 120 bp. All the *T. miscellus* individuals showed *T. pratensis* pattern of expression in terms of heterozygosity and percentages of polymorphic transcripts (~50:50 both C and T alleles) (Table 4.4).

4.4.2.3 Expression of C-class genes

Along with B-function genes, C-class genes specify stamen identity in the third whorl and C-class genes alone specify carpel identity (Krizek & Fletcher, 2005). *Tragopogon* transcripts orthologous to sunflower C-class genes: *HAM45*, *HAM59 and HaAG* (Dezar *et al.*, 2003; Shul'ga *et al.*, 2008) had four, five and two SNPs, respectively (*HAM45*: 377 bp, 458 bp, 485 bp and 529 bp; *HAM59*: 507 bp, 634 bp, 639 bp, 675 bp, and 699 bp; *HaAG*: 368 bp and 408 bp). *Tragopogon* orthologs from *T. miscellus* polyploids showed almost an equal expression of polymorphic transcripts for both *HAM45* and *HAM59* at all the SNPs for all the *T. miscellus* individuals (Table 4.4). *Tragopogon* orthologs of *HaAG* carried 'A' and 'G' variants at 368 bp in both parents with 'A' variant most frequent in parents (70-90%) but in the *T. miscellus* polyploids 'G' variant was frequent (66-83%) except natural short liguled (2604-22). At another SNP (C/T) at 408 bp both parents and *T. miscellus* polyploids carried 'T' allele more frequently (64-91%) than 'C' allele (9-36%) except natural short liguled (2604-22) which equally expressed both variants at 408bp (Table 4.4). *Tragopogon* orthologs of the *Chrysanthemum* C-class gene *CDM37* and orthologs of *Gerbera hybrida* C-class gene *gaga1* had variability but did not have polymorphisms between parents, while orthologs of *gaga2* had two SNPs at 308 bp and 677 bp. All the natural and synthetic *T. miscellus* polyploids equally expressed those variants (Table 4.4). There are also other C-class genes reported in *Gerbera hybrida* such as *Gerbera regulator of capitulum development 1* and *2* (*GRCD1 and GRCD2*). *Tragopogon* orthologs of *GRCD2* had lots of variability but there were no any polymorphisms detected between parents, while orthologs of *GRCD1* had one SNP at 169 bp. In *T. pratensis* and all the *T. miscellus* polyploids transcripts carrying 'T' at that SNP were more frequent compared to *T. dubius* where transcripts carrying 'C' were more common (Table 4.4).

4.4.2.4 Expression of E-class genes

E-class genes involve *SEPALLATA*-like genes that are required at all stages in floral development in *Arabidopsis* and *Petunia* (Pelaz *et al.*, 2000; Vandenbussche *et al.*, 2003). In sunflower, *HAM137* performs E function. *Tragopogon* transcripts orthologous to *HAM137* had two SNPs at 135 bp and 428 bp. All the *T. miscellus* polyploids abundantly expressed the transcripts that were frequently expressed in *T. pratensis* (T-77% at 135 bp, A-75% at 428 bp) (Table 4.4). E-function in *Chysanthemum* is performed by *CDM44*. A small number of *Tragopogon* transcripts mapped to this gene with no polymorphisms between the diploids.

4.4.2.5 Floral symmetry genes

Expression of the floral symmetry genes was determined in the *Tragopogon* diploids (*T. dubius* and *T. pratensis*) and *T. miscellus* polyploids by mapping *Tragopogon* transcriptome against *CYCLOIDEA*-like genes which has been already reported in the members of Asteraceae (Sunflower and *Gerbera hybrida*). Very few *Tragopogon* transcripts mapped to these *CYC*-like genes. Some *Tragopogon* transcripts orthologous to Sunflower *CYC*-like gene *HaCYC2b* (Chapman *et al.*, 2008) had one SNP at 382 bp. Both natural and synthetic long-liguled individuals equally expressed polymorphic transcripts, while the natural short-liguled *T. miscellus* individual (2604-22) frequently expressed transcripts with 'T' variant (75%) similar to *T. pratensis* (71%) and synthetic short-liguled *T. miscellus* individual (111-4) had more transcripts with 'C' variant (80%) like *T. dubius* (93%) (Table 4.4). Some *Tragopogon* transcripts also mapped to *CYC*-like gene *ghCyc4* reported in *Gerbera hybrida* (Broholm *et al.*, 2008). *Tragopogon* orthologs carried two SNPs at 399 bp and 421 bp. Both natural *T. miscellus* polyploids followed *T. dubius* expression pattern in being homozygous while synthetic long-liguled individual (129-7) expressed *T. pratensis*-specific homeolog at higher level (76%). *Tragopogon* transcripts orthologous to another *Gerbera CYC*-like gene *ghCyc2* had one SNP at

357 bp (Table 4.4). Both natural and synthetic long-liguled individuals (2605-9 and 129-7, respectively) showed higher expression of *T. pratensis*-specific homeolog, while natural and synthetic short-liguled individuals (2604-22 and 111-4) expressed other polymorphic allele at a high frequency.

Expression of the other flower development genes such as *LEAFY* and *FLOWERING TIME LOCUS* genes, which are involved in the flowering induction in many plant species, were also analyzed in *Tragopogon* diploids and polyploids. None of the transcripts mapped to previously reported *LEAFY* gene, while few transcripts mapped to *FLOWERING TIME LOCUS* genes *LsFT* reported in *Lactuca sativa* (Fukuda *et al.*, 2011). There were three SNPs in *Tragopogon* orthologs at 1920 bp, 1967 bp and 1984 bp. All the *T. miscellus* individuals equally expressed both polymorphic transcripts except synthetic short-liguled (111-4) which followed *T. pratensis* pattern of expression in terms of percentages of polymorphic transcripts (Table 4.4). None of the *Tragopogon* transcripts mapped to Sunflower flowering locus gene-*Flowering locus T2* (Blackman *et al.*, 2011).

4.5 Discussion

The presence of genetic and epigenetic variation between reciprocal hybrids and the effects of that variation on morphological characters have often been attributed to cytoplasmic inheritance, genomic imprinting or maternal effects (Birky, 1995; Miko, 2008; He *et al.*, 2010). Correns (1909) showed that the development of different leaf colors in reciprocal hybrids of *Mirabilis jalapa* was caused by uniparental inheritance of leaf color. Similarly, the formation of different inflorescence morphologies in reciprocally formed (short- and long-liguled) *Tragopogon miscellus* polyploids has been prourposed to result from a maternal influence. To test this hypothesis, transcriptome data for floral identity genes was analyzed from the inflorescences of diploid parents (*T. dubius* and *T. pratensis*) and their reciprocal *T. miscellus* polyploids in this study.

4.5.1 Divergence among parental species

Genetic variation in the natural populations of *T. dubius* and *T. pratensis* has been well documented previously. *T. dubius* has been reported to be more variable compared to *T. pratensis* (Soltis *et al.*, 1995a; Symonds *et al.*, 2010). Considerable variation has also been detected within both parental species for the floral identity genes studied here. Surprisingly *T. pratensis* was more heterozygous compared to *T. dubius*, which was mostly found to be homozygous and less variable in contrast to previous findings (Soltis *et al.*, 1995a; Mavrodiev *et al.*, 2005; Symonds *et al.*, 2010). Since only a small number of genes were analyzed here, analysis of the remaining transcriptome data might shed more light on this level of divergence between parental species. Similarly, only one individual of each diploid was sequenced and these might not accurately reflect the variation present within the species in the Palouse.

4.5.2 Transcript abundance or expression of floral identity genes

For the first time we report orthologs of the floral identity genes in *Tragopogon* species. Because the analysis of progenitor-specific transcript abundance was complicated by the fact that divergence between parental species was not informative, we were unable to investigate the maternal influence on the formation of short- and long-liguled floral forms in this preliminary study. However, a considerable number of *Tragopogon* transcripts mapped to floral identity genes from other species suggesting that MADS box genes were expressed well at the time of harvesting those inflorescences for generating transcriptome data.

4.5.2.1 Transcript abundance for A-class genes

Tragopogon transcripts orthologous to floral identity genes from Sunflower (*HAM75*), *Chrysanthemum* (*CDM111*) and *Gerbera hybrida* (*gsqua1*) from *APETALA1* and *SQUAMOSA* subfamilies of class A mapped abundantly (Table 3) (Yu *et al.*, 1999; Shchennikova *et al.*, 2004; Shul'ga *et al.*, 2008). From the transcript abundance, among other A-class genes

CDM111 seemed to be most abundantly expressed in *Tragopogon* species (Table 3). *CDM111* has been reported to be predominantly expressed in inflorescence bracts and petals of both ray and disk florets in *Chrysanthemum* (Shchennikova *et al.*, 2004) like other *SQUAMOSA* genes (Huijser *et al.*, 1992). Transcript abundance of this gene in *Tragopogon* should result from its expression in the petals of ray and disk florets only because young and mature floral buds were harvested without bracts for RNA extraction. Moreover, higher expression of this transcription factor may be related to harvesting stage of *Tragopogon* inflorescences as the expression of these MADS box proteins is highly dependent on the floral developmental stages (Ng & Yanofsky, 2001). Nothing could be inferred in terms of progenitor-specific expression for any of the class A orthologs, as these were heterozygous in *T. pratensis* and carried almost an equal proportion of transcripts with these two alleles in the polyploids.

4.5.2.2 Transcript abundance for B-class genes

Tragopogon transcripts mapped abundantly to *HAM31* among other sunflower B-class genes, *CDM86* among other B-class genes in *Chrysanthemum* and *gglo1* in *Gerbera hybrida* (Table 4.3) (Yu *et al.*, 1999; Shchennikova *et al.*, 2004; Shul'ga *et al.*, 2008). All those genes belong to *PISTILLATA*/*GLOBOSA* subfamilies, while other B-class genes in those species also mapped to *APETALA3*/*DEFICIENS* subfamilies but at a lower level (Table 4.3). This suggests that *AP3*/*DEF* subfamilies were not highly expressed at the time of inflorescence collection. B-class genes are responsible for 2nd (petal) and 3rd (stamen) whorl development (Krizek & Fletcher, 2005). All these genes showing higher transcript abundance have been reported to be expressed in these whorls in their respective species, for instance, *gglo1* in *Gerbera hybrida* was reported to be expressed in the stamens (3rd whorl) of developing trans florets (trans florets are intermediate florets forming a whorl that surrounds central disk florets of the capitulum and trans florets are themselves surrounded by a whorl of ray florets in *Gerbera hybrida*) and disc florets of the capitulum (Yu *et al.*, 1999; Shchennikova *et al.*, 2004; Shul'ga *et*

al., 2008). In accordance with the expression of above genes, transcript abundance of these B-class genes in *Tragopogon* might be associated with their spatial expression in the stamens of ray and disc florets.

Although progenitor-specific expression could not be specified to these B-class orthologs, some expression differences based on *T. pratensis*-specific homeolog could be assigned as *T. pratensis* homeolog was heterozygous and carried ~ 50% transcripts which were not present in *T. dubius* (as mentioned in results). So, expression of those *T. pratensis*-specific transcripts in *T. miscellus* polyploids could be discussed here. For instance, natural long-liguled *T. miscellus* (2605-9) which has *T. dubius* as maternal parent (Soltis & Soltis, 1989) showed lower expression of *HAM91* orthologs from *T. pratensis* (27-33%) while other transcripts common between both parents were expressed at a high rate (67-73%) at four SNPs (Table 4.4). Similarly, synthetic short-liguled *T. miscellus* (111-4) which has *T. pratensis* as maternal parent (Tate *et al.*, 2009b) showed higher abundance of *gdef2* orthologs from *T. pratensis*-specific transcripts (81%) at first two SNPs. It could be assumed that the differential expression of the *T. pratensis* homeolog observed here might be cause by maternal influence (Table 4.4).

4.5.2.3 Transcript abundance for C- and E-class genes

Higher abundance of *Tragopogon* transcripts was observed for the orthologs of *HAM59* (sunflower), *CDM37* (*Chrysanthemum*) and both *gaga1* and *gaga2* (*Gerbera hybrida*) (Table 4.3). All these genes belong to the *AGAMOUS* subfamily. Class C genes are generally involved in the identity of 3rd (stamen) and 4th (carpel) whorl (Krizek & Fletcher, 2005). Shulga et al. (2008) reported the expression of *HAM59* in the stamens and carpel of tubular/disc florets, but its expression was absent in the ray florets (where stamens abort) of sunflower suggesting its role in the development of stamens. Similarly, transcript abundance

of *Tragopogon* orthologs could be assumed to follow the above pattern of expression. Among E-class genes, *Tragopgon* orthologs mapped well to sunflower E-class gene *HAM137* from *SEPALLATA3* subfamily compared to other genes (*CDM44*) from *Chrysanthemum,* suggesting *Tragopogon* orthologs for class E genes are more closely related to *HAM137* (Table 4.3).

4.5.2.4 Floral symmetry genes

Floral symmetry (zygomorphic and actinomorphic) has been reported to be controlled by a family of *CYCLOIDEA*-like genes (Hileman & Cubas, 2009). In addition to determining bilateral symmetry (Zygomorphic) (Endress, 1998), the role of *CYC*-like genes in developing novel floral morphologies has been widely reported. For instance, Song et al. (2009) reported the contribution of these genes in the abortion of ventral stamens in the flowers of *Opithandra* (Gesneriaceae). These *CYC*-like genes have also been studied in the members of Asteraceae including sunflower and *Gerbera hybrida,* where *CYC*-like genes have been reported to be expressed only in ray florets (zygomorphic), but not in the disc florets (actinomorphic), suggesting their role in the formation of diverse floret types in the complex inflorescence of Asteraceae (Broholm *et al.*, 2008; Chapman *et al.*, 2008). Expression of this gene family has not been reported previously in *Tragopogon* and in the present study we found very few orthologous transcripts mapping to these *CYC*-like genes reported in both sunflower and *Gerbera hybrid* (Table 4.3). This again suggests temporal expression pattern of these genes which may not have been expressed at the time of *Tragopogon* inflorescence collection.

We were specifically interested in *CYC*-like gene expression because previously they have been reported to control ligule length in the ray and disk florets in *Gerbera hybrida* (Broholm *et al.*, 2008). Broholm et al. (2008) reported the over-expression of *GhCyc2* genes was related to longer petals in disk and trans florets, while its suppression was associated with short ligule length in trans florets of *Gerbera hybrida*. As reciprocal *T. miscellus* polyploids studied here also have different floral structures in terms of length of ligules, we

were expecting that these genes might be involved in determining maternal influence in these polyploids. However, due to an absence of progenitor-specific variation between parents (*T. dubius* and *T. pratensis*), it was not possible to infer homeologous expression in *T. miscellus* polyploids for the above floral symmetry genes and floral identity MADS box genes discussed above.

Nonetheless, this preliminary study identified orthologs of MADS-box transcription factors and other floral developmental genes. A more comprehensive analysis of the resulting transcriptomic data generated from the inflorescences of the synthetic and natural short- and long-liguled *T. miscellus* polyploids would yield valuable findings. Moreover, a thorough developmental study of the individual floral whorls and inflorescences (using morphological and transcriptomic analysis) would be required to identify the genes responsible for these differing inflorescence forms.

Table 4.3 Total read count for *Tragopogon* transcripts mapping to ABC genes for each of the diploid parents and *T. miscellus* polyploid.

		Read count for HiSeq samples (100bp long reads)						**Read count for MiSeq samples (250bp long reads)**					
Gene	Orthologs in *Arabidopsis/Antirhinum*	*T.dubius*	*T. pratensis*	*T. miscellus*				*T.dubius*	*T. pratensis*	*T. miscellus*			
				2609-22	**2604-22**	**129-7**	**111-4**			**2609-22**	**2604-22**	**129-7**	**111-4**
HAM75	*API* and *SQUA*	305	317	288	96	111	642	48	63	46	67	21	69
HAM92	*API* and *SQUA*	166	179	173	58	86	397	13	26	11	26	3	33
CDM111	*API* and *SQUA*	671	748	692	278	270	1673	89	102	78	130	23	123
CDM8	*API* and *SQUA*	0	0	1	0	1	2	0	0	0	0	0	0
CDM41	*API* and *SQUA*	85	110	158	32	69	170	14	22	20	9	1	18
gsqua1	*SQUA*	588	452	438	164	574	966	8	1	0	2	1	2
Orthologs of B-class genes													
HAM31	*PI and GLO*	1218	1382	648	584	1734	1414	162	137	82	140	152	147
HAM2	*AP3 and DEF*	482	431	174	224	1327	638	71	61	23	53	115	57
HAM63	*AP3 and DEF*	59	20	21	12	43	15	5	1	2	2	5	1
HAM91	*AP3 and DEF*	99	104	198	33	132	130	10	6	17	5	6	8
HaAP3	*AP3*	0	5	0	0	3	3	0	0	0	0	0	0
CDM86	*PI*	3830	4023	2183	1491	6574	5427	370	364	167	322	369	449
CDM19	*AP3*	901	703	287	313	2116	1135	139	131	35	107	195	109
CDM115	*AP3*	21	17	11	6	43	17	3	2	0	1	0	0
gglo1	*GLO*	4492	4157	2504	1000	7775	6048	640	470	317	495	693	593
gdef1	*DEF*	200	249	364	72	284	230	11	21	22	12	17	16
gdef2	*DEF*	807	604	574	314	1748	1172	108	135	76	88	159	100
Lac-Lsa	*PI*	889	1101	564	380	1379	1236	113	80	48	77	91	97
Orthologs of C-class genes													
HAM45	*AGAMOUS* and *PLENA*	180	119	132	40	311	279	34	17	12	19	33	25
HAM59	*AGAMOUS* and *PLENA*	527	362	212	163	610	658	82	51	26	58	50	58
HaAG	*AGAMOUS*	261	198	173	72	534	482	9	29	7	12	26	27
CDM37	*AGAMOUS*	434	268	217	69	691	568	48	47	23	42	44	38

gaga1	*AGAMOUS*	457	340	230	97	716	759	54	39	23	41	42	61
gaga2	*AGAMOUS*	312	300	152	97	504	611	41	44	21	27	33	32
GRCD1	*Gerbera regulater of Capitulum development*	396	1069	468	412	1650	1415	73	256	43	130	145	151
GRCD2	*Gerbera regulater of Capitulum development*	434	370	375	125	689	754	45	78	27	58	44	55
Orthologs of E-class genes													
HAM137	*SEP3 or AGL9*	570	972	408	458	1565	1307	84	161	49	96	113	128
CDM44	*SEP3 or AGL2*	19	24	28	8	37	50	0	1	1	0	0	0
Orthologs of floral symmetry genes													
HaCyc2b	*Cycloidea-like*	17	8	5	4	32	35	1	0	0	0	0	1
ghCyc2	*Cycloidea-like*	4	3	1	2	8	3	0	0	0	0	0	0
ghCyc4	*Cycloidea-like*	7	6	4	3	15	18	0	1	0	0	0	0
LsFT	*Flowering Time Locus*	12	35	33	15	33	41	0	3	4	1	7	2

Table 4.4 Percentages of parental transcripts in *Tragopogon* diploid and polyploids orthologous to ABC class genes from other groups of Asteraceae.

Orthologs of A-Class genes								
Gene	**Reference species**	**Species**	**SNP 1**	**SNP 2**	**SNP 3**	**SNP 4**	**Position of SNP 5**	**Position of SNP 6**
			Percentage of transcripts	**Percentage of transcripts**	**Percentage of transcripts**	**Percentage of transcripts**	**Percentage of transcripts**	**Percentage of transcripts**
			294bp	319bp	328bp			
HAM75	*Halianthus annuus*	*T. dubius*	A-97% G-3%	C-100%	C-100%			
		T. pratensis	A-44% C-56%	C-44% T-56%	C-40% T-57%			
		2604-22 *T. miscellus*(NS)	A-53% C-48%	C-50% T-50%	C-44% T-56%			
		2605-9 *T. miscellus*(NL)	A-65% C-35%	C-69% T-31%	C-69% T-31%			
		111-4 *T. miscellus*(SS)	A-49% C-49%	C-51% T-49%	C-50% T-50%			
		129-7 *T. miscellus*(SL)	A-58% C-42%	C-58% T-42%	C-55% T-45%			
			382bp	374bp	427bp			
HAM92	*Halianthus annuus*	*T. dubius*	C-100%	C-100%	A-100%			
		T. pratensis	C-73% T-27%	C-67% T-33%	A-63% G-37%			
		2605-9 *T. miscellus*(NL)	C-52% T-48%	C-45% T-55%	A-43% G-57%			
		2604-22 *T. miscellus*(NS)	C-62% T-38%	C-54% T-46%	A-46% G-54%			
		129-7 *T. miscellus*(SL)	C-34% T-66%	C-31% T-69%	A-35% G-63%			
		111-4 *T. miscellus*(SS)	C-52% T-48%	C-50% T-50%	A-47% G-53%			
			433bp	480bp	580bp			
CDM111	*Chrysanthemum*	*T. dubius*	A-100%	G-100%	C-2% T-97%			
		T. pratensis	A-45% G-54%	A-55% G-45%	C-57% T-40%			
		2604-22 *T. miscellus*(NS)	A-54% G-46%	A-50% G-50%	C-44% T-58%			
		2605-9 *T. miscellus*(NL)	A-43% G-57%	A-54% G-46%	C-52% T-48%			
		111-4	A-36% G-64%	A-65% G-35%	C-62% T-37%			

		T. miscellus(SS)						
		129-7 *T. miscellus*(SL)	A-53% G-47%	A-54% G-46%	C-58% T-42%			
			167bp	184bp	325bp			
CDM41	*Chrysanthemum*	*T. dubius*	T-100%	C-100%	G-100%			
		T. pratensis	G-50% T-50%	A-57% C-43%	C-56% G-44%			
		2604-22 *T. miscellus*(NS)	G-60% T-40%	A-64% C-36%	C-62% G-33%			
		2605-9 *T. miscellus*(NL)	G-38% T-62%	A-43% C-57%	C-38% G-63%			
		111-4 *T. miscellus*(SS)	G-25% T-75%	A-25% C-75%	C-60% G-40%			
		129-7 *T. miscellus*(SL)	None of reads mapped	NA	NA			
Orthologs of B Class genes								
			256bp	274bp	343bp			
HAM2	*Halianthus annuus*	*T. dubius*	C-2% T-98%	A-100%	G-100%			
		T. pratensis	G-47% T-53%	A-55% G-45%	A-45% G55%			
		2604-22 *T. miscellus*(NS)	G-54% T-46%	A-46% G-54%	A-59% G41%			
		2605-9 *T. miscellus*(NL)	G-55% T-45%	A-50% G-50%	A-41% G59%			
		111-4 *T. miscellus*(SS)	G-55% T-45%	A-42% G-58%	A-57% G43%			
		129-7 *T. miscellus*(SL)	G-62% T-38%	A-36% G-64%	A-64% G36%			
			496bp					
HAM31	*Halianthus annuus*	*T. dubius*	A-99% C-1%					
		T. pratensis	A-40% C-56%					
		2604-22 *T. miscellus*(NS)	A-40% C-59%					
		2605-9 *T. miscellus*(NL)	A-48% C-50%					
		111-4 *T. miscellus*(SS)	A-57% C-42%					

		129-7 *T. miscellus*(SL)	A-50% C-50%					
			107bp	114bp	154bp	157bp	160bp	
HAM91	*Halianthus annuus*	*T. dubius*	A-100%	A-96% G-4%	G-92% T-8%	A-99% C-1%	G-100%	
		T. pratensis	A-57% C-43%	A-58% G-42%	T-100%	A-58% C-42%	A-41% G-59%	
		2605-9 *T. miscellus*(NL)	A-73% C-27%	A-69% G-31%	T-100%	A-67% C-33%	A-33% G-67%	
		2604-22 *T. miscellus*(NS)	A-60% C-40%	A-63% G-37%	T-100%	A-59% C-41%	A-41% G-59%	
		129-7 *T. miscellus*(SL)	A-52% C-48%	A-47% G-53%	T-100%	A-50% C-50%	A-51% G-49%	
		111-4 *T. miscellus*(SS)	A-64% C-36%	A-58% G-42%	T-100%	A-67% C-33%	A-33% G-67%	
			500bp					
CDM86	*Chrysanthemum*	*T. dubius*	A-100%					
		T. pratensis	A-48% C-52%					
		2605-9 *T. miscellus*(NL)	A-46% C-53%					
		2604-22 *T. miscellus*(NS)	A-48% C-51%					
		129-7 *T. miscellus*(SL)	A-50% C-49%					
		111-4 *T. miscellus*(SS)	A-51% C-48%					
			193bp	461bp	529bp	570bp		
CDM19	*Chrysanthemum*	*T. dubius*	G-100%	A-1% C-99%	G-2% T-98%	C-4% G-96%		
		T. pratensis	A-19% G-81%	A-52% C-47%	C-53% T-47%	A-45% G-55%		
		2604-22 *T. miscellus*(NS)	A-41% G-57%	A-47% C-52%	C-52% T-48%	A-46% G-54%		
		2605-9 *T. miscellus*(NL)	A-57% G-43%	A-47% C-53%	C-50% T-50%	A-0% G-100%		
		111-4 *T. miscellus*(SS)	A-37% G-63%	A-65% C-35%	C-70% T-30%	A-33% G-67%		
		129-7 *T. miscellus*(SL)	A-59% G-41%	A-58% C-42%	C-60% T-40%	A-37% G-63%		
			154bp					

gglo1	*Gerbera hybrida*	*T. dubius*	C-0% T-100%					
		T. pratensis	C-51% T-49%					
		2605-9 *T. miscellus*(NL)	C-51% T-49%					
		2604-22 *T. miscellus*(NS)	C-48% T-52%					
		129-7 *T. miscellus*(SL)	C-47% T-53%					
		111-4 *T. miscellus*(SS)	C-45% T-55%					
			77bp	84bp	124bp	127bp	129bp	504bp
gdef1	*Gerbera hybrida*	*T. dubius*	A-100%	A-100%	G-82% T-12%	A-90% C-5%	G-100%	G-100%
		T. pratensis	A-24% C-76%	A-22% G-77%	T-100%	A-13% C-87%	A-80% G-20%	G-57% T-43%
		2605-9 *T. miscellus*(NL)	A-28% C-71%	A-28% G-72%	T-100%	A-20% C-80%	A-61% G-39%	G-48% T-52%
		2604-22 *T. miscellus*(NS)	A-19% C-81%	A-21% G-79%	T-100%	A-15% C-85%	A-54% G-46%	G-62% T-34%
		129-7 *T. miscellus*(SL)	A-29% C-71%	A-27% G-73%	T-100%	A-19% C-81%	A-83% G-17%	G-44% T-56%
		111-4 *T. miscellus*(SS)	A-35% C-65%	A-33% G-67%	T-100%	A-25% C-75%	A-73% G-27%	G-54% T-45%
			163bp	185bp	275bp			
gdef2	*Gerbera hybrida*	*T. dubius*	T-100%	C-98% T-2%	C-26% T-74%			
		T. pratensis	T-47% C-53%	C-48% A-51%	C-55% T-45%			
		2605-9 *T. miscellus*(NL)	T-45% C-55%	C-44% A-56%	C-99% T-1%			
		2604-22 *T. miscellus*(NS)	T-63% C-38%	C-57% A-41%	C-98% T-2%			
		129-7 *T. miscellus*(SL)	T-44% C-56%	C-43% A-56%	C-56% T-44%			
		111-4 *T. miscellus*(SS)	T-19% C-81%	C-17% A-82%	C-81% T-19%			
			120bp					
Lac-Lsa	*Lactuca sativa*	*T. dubius*	C-0% T-100%					
		T. pratensis	C-50% T-50%					
		2605-9 *T. miscellus*(NL)	C-55% T-45%					

		2604-22 *T. miscellus*(NS)	C-51% T-49%					
		129-7 *T. miscellus*(SL)	C-48% T-52%					
		111-4 *T. miscellus*(SS)	C-46% T-54%					
Orthologs of C Class genes								
			377bp	458bp	485bp	529bp		
HAM45	*Halianthus annuus*	*T. dubius*	C-100%	A-93% G-7%	A-100%	C-92% T-8%		
		T. pratensis	C-53% T-47%	A-76% G-24%	A-70% T-30%	C-44% T-56%		
		2604-22 *T. miscellus*(NS)	C-57% T-43%	A-78% G-22%	A-78% T-22%	C-29% T-71%		
		2605-9 *T. miscellus*(NL)	C-64% T-36%	A-11% G-89%	A-61% T-39%	C-63% T-38%		
		111-4 *T. miscellus*(SS)	C-72% T-28%	A-17% G-83%	A-56% T-44%	C-53% T-47%		
		129-7 *T. miscellus*(SL)	C-58% T-42%	A-21% G-79%	A-63% T-37%	C-35% T-62%		
			507bp	634bp	639bp	675bp	699bp	
HAM59	*Halianthus annuus*	*T. dubius*	C-2% G-98%	C-100%	C-100%	G-100%	A-100%	
		T. pratensis	A-52% G-48%	C-67% T-33%	C-56% T-44%	C-44% G-55%	A-76% G-24%	
		2605-9 *T. miscellus*(NL)	A-46% G-54%	C-72% T-28%	C-60% T-40%	C-31% G-69%	A-93% G-7%	
		2604-22 *T. miscellus*(NS)	A-51% G-49%	C-68% T-32%	C-53% T-47%	C-43% G-57%	A-77% G-23%	
		129-7 *T. miscellus*(SL)	A-50% G-50%	C-61% T-39%	C-56% T-41%	C-35% G-65%	A-75% G-25%	
		111-4 *T. miscellus*(SS)	A-60% G-40%	C-68% T-32%	C-57% T-43%	C-34% G-66%	A-77% G-23%	
			368bp	408bp				
HαAG	*Halianthus annuus*	*T. dubius*	A-90% G-10%	C-36% T-64%				
		T. pratensis	A-69% G-31%	C-32% T-68%				
		2605-9 *T. miscellus*(NL)	A-17% G-83%	C-9% T-91%				
		2604-22 *T. miscellus*(NS)	A-86% G-14%	C-56% T-44%				
		129-7	A-21% G-79%	C-14% T-86%				

		T. miscellus(SL)						
		111-4 *T. miscellus*(SS)	A-34% G-66%	C-23% T-77%				
			308bp	677bp				
gaga2	*Gerbera hybrida*	*T. dubius*	G-100%	T-100%				
		T. pratensis	A-51% G-49%	C-59% T-41%				
		2605-9 *T. miscellus*(NL)	A-31% G-69%	C-57% T-43%				
		2604-22 *T. miscellus*(NS)	A-60% G-40%	C-68% T-32%				
		129-7 *T. miscellus*(SL)	A-44% G-56%	C-66% T-34%				
		111-4 *T. miscellus*(SS)	A-44% G-56%	C-60% T-40%				
			169bp					
GRCD1	*Gerbera hybrida*	*T. dubius*	C-99% T-1%					
		T. pratensis	C-16% T-83%					
		2605-9 *T. miscellus*(NL)	C-22% T-78%					
		2604-22 *T. miscellus*(NS)	C-17% T-83%					
		129-7 *T. miscellus*(SL)	C-22% T-78%					
		111-4 *T. miscellus*(SS)	C-19% T-81%					
Orthologs of E Class genes								
			135bp	428bp				
HAM137	*Halianthus annuus*	*T. dubius*	C-91% T-9%	A-31% C-69%				
		T. pratensis	C-23% T-77%	A-75% C-25%				
		2605-9 *T. miscellus*(NL)	C-31% T-69%	A-100%				
		2604-22 *T. miscellus*(NS)	C-23% T-77%	A-100%				
		129-7 *T. miscellus*(SL)	C-28% T-71%	A-64% C-36%				
		111-4 *T. miscellus*(SS)	C-27% T-73%	A-35% C-65%				

Orthologs of Floral Symmetry Genes								
			382bp					
HaCyc2b	*Helianthus annuus*	*T. dubius*	C-93% T-7%					
		T. pratensis	C-29% T-71%					
		2605-9 *T. miscellus*(NL)	C-50% T-50%					
		2604-22 *T. miscellus*(NS)	C-25% T-75%					
		129-7 *T. miscellus*(SL)	C-63% T-38%					
		111-4 *T. miscellus*(SS)	C-80% T-20%					
			357bp					
ghCyc2	*Gerbera hybrida*	*T. dubius*	A-100%					
		T. pratensis	A-33% G-67%					
		2605-9 *T. miscellus*(NL)	G-100%					
		2604-22 *T. miscellus*(NS)	A-100%					
		129-7 *T. miscellus*(SL)	A-38% G-63%					
		111-4 *T. miscellus*(SS)	A-67% G-33%					
			399bp	421bp				
ghCyc4	*Gerbera hybrida*	*T. dubius*	T-100%	C-100%				
		T. pratensis	A-50% T-50%	T-40% C-60%				
		2605-9 *T. miscellus*(NL)	T-100%	C-100%				
		2604-22 *T. miscellus*(NS)	T-100%	C-100%				
		129-7 *T. miscellus*(SL)	A-76% T-24%	T-76% C-24%				
		111-4 *T. miscellus*(SS)	A-24% T-76%	T-24% C-76%				

			1920bp	1967bp	1984bp			
LsFT	*Lactuca sativa*	*T. dubius*	C-100%	C-11% T-89%	A-75% C-25%			
		T. pratensis	C-9% T-91%	C-88% T-12%	A-100%			
		2605-9 *T. miscellus*(NL)	C-40% T-60%	C-54% T-46%	A-100%			
		2604-22 *T. miscellus*(NS)	C-50% T-50%	C-44% T-56%	A-100%			
		129-7 *T. miscellus*(SL)	C-35% T-65%	C-66% T-34%	A-100%			
		111-4 *T. miscellus*(SS)	C-7% T-93%	C-89% T-11%	A-97% C-3%			

Note: NS, NL, SS and SL correspond to Natural Short-liguled, Natural Long-liguled, Synthetic Short liguled and Synthetic Long-liguled respectively.

CHAPTER FIVE

This chapter is in preparation as:

Sehrish T, *Symonds VV, Soltis DE, Soltis PS, and Tate JA . Gene silencing via DNA methylation in naturally occurring Tragopogon miscellus (Asteraceae) allopolyploids. In prep. for Molecular Biology and Evolution.*

5. Gene silencing via DNA methylation in naturally occurring *Tragopogon miscellus* (Asteraceae) allopolyploids

5.1 Abstract

Hybridization coupled with whole-genome duplication (allopolyploidy) leads to a variety of genetic and epigenetic modifications in the resultant merged genomes. In particular, gene loss and gene silencing are commonly observed post-polyploidization. Here, we investigated DNA methylation as a potential mechanism for gene silencing in *Tragopogon miscellus*, a recent and recurrently formed allopolyploid. This species, which also exhibits extensive gene loss, was formed from the diploids *T. dubius* and *T. pratensis*. Comparative bisulfite sequencing revealed methylation of CpG sites in the parental homeologs of two loci (S2 and TDF-44) that were previously identified as silenced in *T. miscellus* individuals relative to the diploid progenitors. The other three loci (S3, S8 and S18) examined did not show methylation, indicating that other transcriptional and post-transcriptional mechanisms are likely responsible for silencing those homeologous loci. These results indicate that polyploids employ diverse mechanisms to respond to the potential shock of genome merger and doubling.

5.2 Introduction

Whole-genome duplication (polyploidy) has played a major role in eukaryotic evolution (Sankoff *et al.*, 2000; Wolfe, 2001; Paterson *et al.*, 2006; Braasch & Postlethwait, 2012; Cañestro, 2012; Hudson & Conant, 2012). In particular, flowering plants have experienced repeated episodes of polyploidy, since they shared a common ancestor with the gymnosperms some 300 million years ago (Soltis *et al.*, 2009; Jiao *et al.*, 2011). Understanding the genomic consequences of polyploidization, particularly when accompanied by hybridization (allopolyploidy), allows insight into the potential for

speciation and adaptation of these novel entities (Ramsey & Schemske, 1998; Madlung, 2013). In particular, the merger and doubling of two divergent genomes can induce different genetic and epigenetic changes in the resulting polyploid (Comai, 2005b; Salmon *et al.*, 2005; Chen, 2007a; Paun *et al.*, 2007; Ma & Gustafson, 2008; Wendel *et al.*, 2012). Genetic modifications can include gene loss, genome down-sizing, variable mutation rates of the duplicated genes (homeologs), chromosomal rearrangements and regulatory incompatibilities resulting from post-transcriptional modifications in the merged genomes (Doyle *et al.*, 2008; Jackson & Chen, 2010; Xiong *et al.*, 2011; Buggs *et al.*, 2012a; Chester *et al.*, 2012; Tang *et al.*, 2012; Wendel *et al.*, 2012; Zielinski & Scheid, 2012; Renny-Byfield *et al.*, 2013). Epigenetic modifications involve heritable changes in gene expression without changes in the nucleotide sequence (Liu & Wendel, 2003; Rapp & Wendel, 2005; Zeng *et al.*, 2012) and may include histone modification, DNA methylation, chromatin remodeling, microRNAs, or prions (Halfmann & Lindquist, 2010; Vanyushin & Ashapkin, 2011; Lee & Shin, 2012). DNA methylation, the addition of a methyl group at position 5 of the pyrimidine ring of cytosine, is a common mechanism associated with gene silencing in polyploids (Finnegan *et al.*, 1998; Chan *et al.*, 2005; Vanyushin, 2006). In general, cytosine methylation is important for maintaining genomic stability and is involved in genomic imprinting, transposon silencing and epigenetic regulation of gene transcription (Martienssen & Colot, 2001; He *et al.*, 2011; Vanyushin & Ashapkin, 2011; Ji & Chen, 2012).

Here, we investigated gene silencing via methylation in the allotetraploid plant *Tragopogon miscellus*. This species formed repeatedly and recurrently during the early 1900s in the western United States, following the introduction of the diploid progenitors, *T. dubius* and *T. pratensis*, from Europe (Ownbey, 1950; Ownbey & Mccollum, 1953; Soltis & Soltis, 1989; Soltis *et al.*, 1995a). Previous studies identified extensive homeolog loss (Tate *et al.*, 2006b; Buggs *et al.*, 2009; Tate *et al.*, 2009a; Buggs *et al.*, 2012a) and chromosomal variation (Chester *et al.*, 2012) in naturally occurring *T. miscellus* populations. Two studies (Tate *et al.*, 2006b; Buggs *et al.*, 2009) also identified

homeologous gene silencing in some individuals of *T. miscellus*. In Tate *et al.* (2006b), the *T. dubius* copy of one locus (TDF-44) was silenced in multiple individuals from Pullman, Washington, and Moscow, Idaho. In Buggs *et al.* (2009), six loci showed variable silencing of *T. dubius* or *T. pratensis* homeologs in a few individuals from five different populations (Oakesdale, Pullman, and Spangle, Washington; Moscow and Garfield, Idaho). In the present study, we used comparative bisulfite sequencing to determine if these loci were silenced by methylation.

5.3 Materials and methods

5.3.1 Plant material

DNA for the diploid parents (*Tragopogon dubius* and *T. prat*ensis) and *Tragopogon miscellus* was the same as used by Tate *et al.* (2006b) and Buggs *et al.* (2009). In all, 19 *T. miscellus* individuals, each of which previously showed gene silencing (TDF-44 in Tate *et al.* 2006; S2, S3, S8, and S18 in Buggs *et al.* 2009; Table 5.1), were examined. Three representatives of each diploid species were also included.

5.3.2 Bisulfite conversion

The basic principle of bisulfite conversion has been explained in the supplementary material. Prior to bisulfite conversion, genomic DNA of the diploid and polyploid samples was digested with *Eco*RV (New England Biolabs, UK), which does not cut within the genes of interest. Two micrograms of genomic DNA were digested in a total volume of 100 µl with 80 units of *Eco*RV, 10X buffer and 10 µg BSA. The reaction was incubated at 37°C overnight (16-18 hours) and the digested DNA cleaned by ethanol precipitation. Bisulfite conversion was carried out using the EZ DNA Methylation kit (Zymo Research, USA). After bisulfite conversion, the single-stranded DNA was quantified using parameters for RNA-40 on a Nanodrop-1000 (Thermo Fisher Scientific, USA).

5.3.3 Amplification and sequencing of genomic and bisulfite-converted DNA

Primers were designed following Warnecke *et al.* (2002). Separate primers were designed to amplify sense and antisense strands as after bisulfite conversion the two strands were not precisely complementary, with additional primers designed to perform nested PCR, using Methyl Primer Express software v. 1.0 (Applied Biosystems, USA). Primers 26-29 bp in length were designed to generate an amplicon of ~300 bp and with a C or T near the 3' end to avoid non-specific binding in the bisulfite-converted DNA. The primers used for amplification of genomic DNA and bisulfite-converted DNA are listed in Table S5.1.

Amplification of bisulfite-converted DNA for the primary PCR reaction was conducted in a total volume of 25 µl with 10 ng template DNA, 10 µM of both gene-specific forward and reverse primers, 10X PCR buffer, 10 mM dNTPs and 1 unit of Takara Ex Taq™ polymerase (Takara Biotechnology, Japan). Genomic and bisulfite-converted DNA was amplified using the following PCR program: 95°C for 5 min, 95°C for 1 min, 53°C for 1 min, 72°C for 1 min for the first 5 cycles, then 44 cycles with 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. Using the nested primers, another PCR was performed using the primary PCR product as template. The resulting nested PCR products were run on a 1.5% agarose gel stained with ethidium bromide and examined using a Gel Doc 2000 system (Bio-Rad, UK). For sequencing, PCR products were treated with Exonuclease I (5U) and Shrimp alkaline phosphatase (0.5 unit) prior to the cycle sequencing reaction using BigDye Terminator v. 3.1 (Applied Biosystems). The purified products were sequenced with both forward and reverse primers on an ABI DNA Analyzer 3770 at Massey Genome Service (Palmerston North, New Zealand). The resulting sequences were assembled and analyzed in Sequencher v. 4.10.1.

5.3.4 Genome walking

In order to determine the methylation status of the promoter region of the five genes, 5'genome walking was performed following the GenomeWalker manual (Clontech

Laboratories) (Siebert *et al.*, 1995). For each gene, two gene-specific reverse primers were designed near the 5' end (Table S5.1). Long and short oligos (NA44 and NA45) to form an adapter to bind to the digested genomic DNA and one of those oligos to act as forward adapter-specific primers (AP1 and AP2) were designed (Table S5.1). The products of each genome walking reaction were cloned using the Invitrogen TOPO TA Cloning Kit (Life Technologies, USA). For detailed protocols see Supplementary Materials and Methods.

5.4 Results and Discussion

Genomic and bisulfite-converted sequences were acquired for five loci [TDF-44 (Tate *et al.*, 2006b) and S2, S3, S8, and S18 (Buggs *et al.*, 2009)] from allopolyploid *Tragopogon miscellus* and the diploid parents *T. dubius* and *T. pratensis* (Table 5.1). Inspection of the promoter and coding regions identified CpG islands, which are common methylation sites in plants (Shawn *et al.*, 2008; Julie & Steven, 2010). The integrity of bisulfite conversion was determined from the conversion of all the Cs that were not adjacent to a G into Ts. The loci studied here all showed complete bisulfite conversion in the genic regions, while incomplete conversion at a few sites was detected in the promoter regions of S8 and TDF-44. All *T. miscellus* individuals showed incomplete conversion in a portion of the promoter region for S8, while only three polyploid individuals (2604-4, 2604-35 and 2605-14) showed incomplete conversion for TDF-44. Given that the majority of the promoter sequence and genic regions were properly converted, the incomplete conversion in these areas for S8 and TDF-44 does not influence the overall interpretation of the results. This low frequency of partial bisulfite conversion as observed here is commonly due to reaction temperature (Grunau *et al.*, 2001; Genereux *et al.*, 2008).

To determine if a parental homeolog was silenced by methylation in the *T. miscellus* individuals, we took advantage of single nucleotide polymorphisms (SNPs) between the diploids that discriminate between the parental copies in the allopolyploid (Fig. 5.1a). CpG methylation of both sense and antisense strands was detected in the genic

and promoter regions of S2 and TDF-44 with their sequences covering 10 and 11 CpG sites, respectively. In the case of TDF-44 (putative leucine-rich repeat transmembrane protein kinase), the *T. dubius* homeolog was methylated in 11 of 12 *T. miscellus* individuals from Pullman and Moscow (Fig. 5.1b, Table 5.1), which confirms the mechanism of silencing observed in Tate *et al.* (2006).

Table 5. 1 Individual plants used in the study and their methylation status for the genes studied; silencing data from Tate *et al.* (2006) and Buggs *et al.* (2009).

Population	Species	Lineage	Locus silenced	Methylated?
Pullman	*T. dubius*	2613-1	NA	
	T. dubius	2613-11	NA	
	T. miscellus	2605-4	$TDF44_d$	Yes
	T. miscellus	2605-7	$TDF44_d$	Yes
	T. miscellus	2605-13	$TDF44_d$	Yes
	T. miscellus	2605-24	$TDF44_d$	Yes
	T. miscellus	2605-28	$TDF44_d$	Yes
	T. miscellus	2605-46	$TDF44_d$	Yes
Moscow	*T. pratensis*	2608-31	NA	
	T. pratensis	2608-35x	NA	
	T. miscellus	2604-4	$TDF44_d$	Yes
	T. miscellus	2604-11	$TDF44_d$	Yes
	T. miscellus	2604-15	$TDF44_d$	Yes
	T. miscellus	2604-22	**T. pratensis* genomic copy lost	
	T. miscellus	2604-24	$TDF44_d$	Yes
	T. miscellus	2604-35	$TDF44_d$	Yes
Spangle	*T. miscellus*	2693-7	$S3_p$	No
	T. miscellus	2693-8	$S3_p$	No
	T. miscellus	2693-14	$S3_p$ $S18_d$	No No
Garfield	*T. dubius*	2687-11	NA	
	T. pratensis	2689-17	NA	
	T. miscellus	2688-3	$S2_d$ $S3_d$ $S8_d$ $S18_d$	Yes No No No
Oakesdale	*T. miscellus*	2671-2	$S8_p$	No
	T. miscellus	2671-8	$S8_p$	No
	T. miscellus	2671-11	$S2_d$	Yes

Note: NA = Not applicable

The exception was individual 2604-22, which retained only the *T. dubius* genomic homeolog and therefore expressed that copy (Tate *et al.*, 2006b). Similarly, we confirm methylation of the S2 (putative RNA binding protein), which was shown to be silenced by Buggs *et al.* (2009) in two *T. miscellus* individuals (one each from Garfield and Oakesdale). In this individual, both parental homeologs showed CpG methylation with more *T. dubius*

copies silenced compared to *T. pratensis* (as determined by the the cloned sequences) resulting in differential expression of parental homeologs in *T. miscellus* polyploids. This result suggests that methylation could regulate the level of expression of parental copies rather than complete silencing.

Analysis of the promoter and genic regions of the other three loci (S3-putative NADP/FAD oxidoreductase, S8-putative Acetyl transferase and S18-porphyrin-oxidoreductase) did not show methylation of any of the CpG sites (Fig 5.1c, Table 5.1; S3, S8 and S18 included 9, 15 and 22 CpG sites, respectively, in genic and promoter regions). Thus, there may be mechanisms other than DNA methylation that are responsible for homeolog specific silencing. For example, histone deacetylation (causing chromatin condensation) is thought to be responsible for transcriptional repression (Kim *et al.*, 2012; Luo *et al.*, 2012; Ma *et al.*, 2013). RNA interference (RNAi) is also widely associated with post-transcriptional silencing via a number of different mechanisms, including mRNA degradation, translational inhibition and the repression of transcription elongation (Carthew & Sontheimer, 2009; Guang *et al.*, 2010; Ketting, 2011; Feng & Guang, 2013). Hence, like other polyploid species (*Spartina*, Salmon *et al.*, 2005; *Arabidopsis*, Zhang *et al.*, 2006; wheat, Hu *et al.*, 2013; rice, Wang *et al.*, 2013b; *Brassica*, Zhang *et al.*, 2013), genome evolution in *Tragopogon miscellus* includes DNA methylation as a mechanism to regulate duplicate gene expression, which we demonstrate here for the first time. Both loci confirmed to be silenced via methylation had the *T. dubius* copy silenced, which, although a small number, may indicate a 'preference' for silencing loci of one progenitor's genome. This result is true of the *T. miscellus* polyploids formed with either *T. dubius* (Pullman) or *T. pratensis* (Garfield, Moscow, Oakesdale, Spangle) as the maternal parent, so there does not seem to be a maternal 'imprinting' influence for the number of loci studied here.

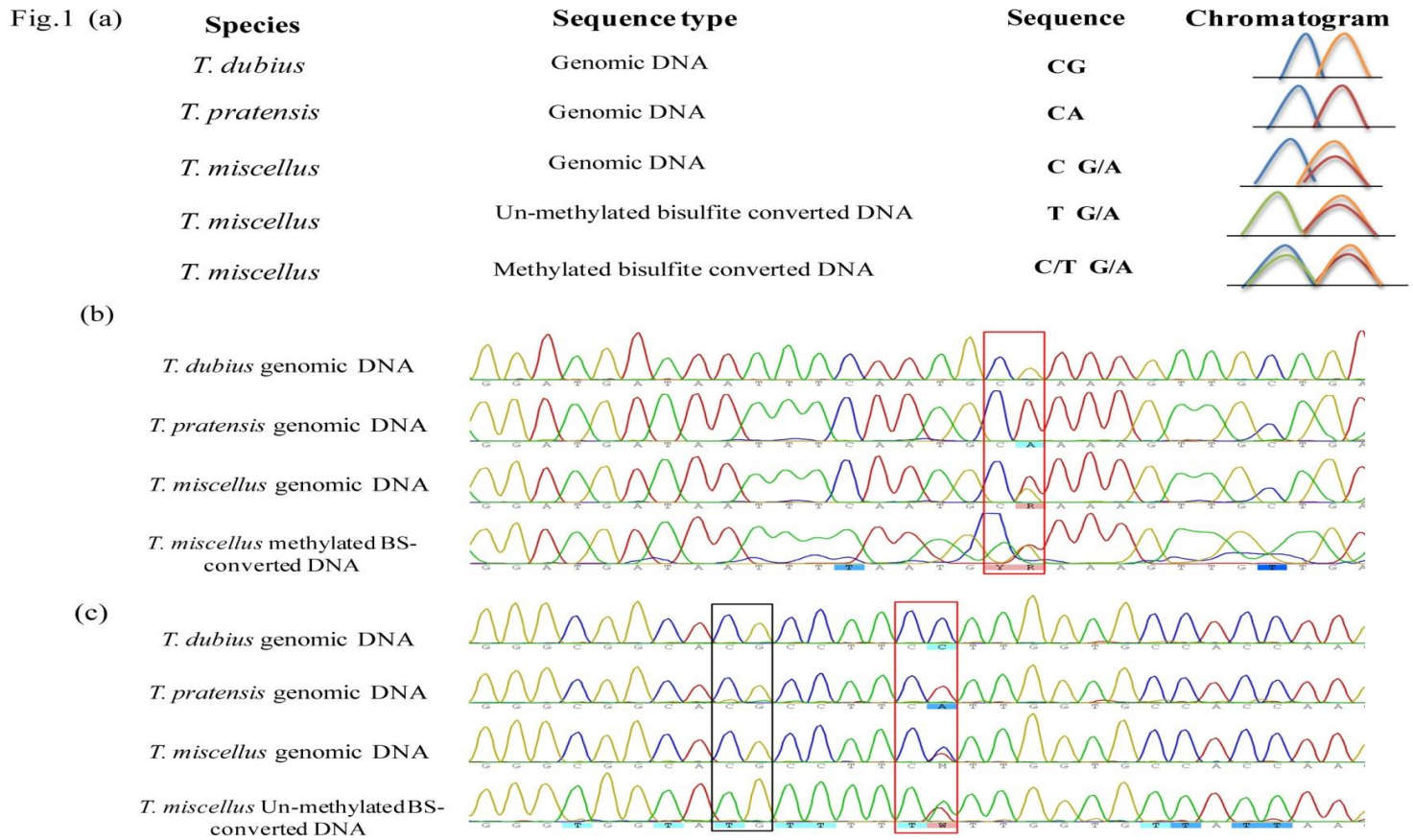


Fig. 5. 1Sequence polymorphisms between the diploid parents (*Tragopogon dubius* and *T. pratensis*) were used to determine if there is homeolog-specific silencing in *T. miscellus* allopolyploids. (a) Diagrammatic illustration of the expected chromatogram peaks for genomic and bisulfite-converted sequences when un-methylated or methylated in allopolyploid *T. miscellus*. This example shows silencing of the *T. dubius* homeolog. (b) Chromatograms of TDF-44 indicating the position of a methylated CpG adjacent to a polymorphic site (red box) in *T. miscellus* compared to the diploids. (c) Chromatograms from S18 showing an un-methylated CpG site in *T. miscellus* (black box) and the location of a polymorphic site between parental copies (red box). Red, blue, green and yellow colors of the chromatogram correspond to A, C, T and G, respectively.BS-converted = bisulfite-converted.

This interpretation is in line with previous studies that have reported a greater tendency of homeolog loss of the *T. dubius* copy compared to *T. pratensis* (Kovarik *et al.*, 2005; Tate *et al.*, 2006b; Matyasek *et al.*, 2007; Buggs *et al.*, 2009; Tate *et al.*, 2009a; Buggs *et al.*, 2010b; Buggs *et al.*, 2012a). Curiously, in the case of rDNA, although *T. dubius* homeologs are more frequently lost from the polyploid genomes, transcription rates of remaining *T. dubius* copies are higher than *T. pratensis* copies (Matyasek *et al.*, 2007). As *T. miscellus* has shown a high frequency of homeolog loss, but little gene silencing based on the studies to date (Tate *et al.*, 2006b; Buggs *et al.*, 2009; Tate *et al.*, 2009a; Buggs *et al.*, 2011; Buggs *et al.*, 2012a), a more comprehensive genome-wide analysis of methylation would help to determine the role of this epigenetic mechanism in shaping the evolution of *Tragopogon* allopolyploid genomes.

5.5 Supplementary material and methods

5.5.1 Principle of bisulfite conversion

DNA methylation does not change the sequence of DNA and is lost after amplification of the DNA fragment containing it. This fact makes the analysis of DNA methylation difficult. Bisulfite conversion is a process that converts unmethylated cytosine into uracil while methylated cytosine remain unchanged. First step is the sulphonation which involves the addition of a sulphonate group (SO_3^-) and one hydrogen ion (H^+) into cytosine ring to form cytosine sulphonate, the second step is the hydrolytic de-amination involving removal of the amine group (NH_4^+) from cytosine ring and the third step is the alkaline de-sulphonation involving removal of the sulphonate group leaving a ring structure corresponding to uracil (Clark *et al.*, 1994). Subsequent amplification of this bisulfite-converted DNA detects uracil as thymine and leaves methylation-specific single nucleotide polymorphism in the sequence of bisulfite converted DNA. This polymorphism is detected by aligning this bisulfite converted sequence with untreated genomic DNA sequence (Smith *et al.*, 2009).

5.5.2 Cloning of BS-converted sequences

Since both parental homeologs in *T. miscellus* polyploids showed CpG methylation of S2 locus, cloning was done to distinguish both parental copies and confirm their methylation. PCR products of BS-converted DNA were cloned from 2671-11 and 2688-3 using the TOPO TA cloning kit (Invitrogen, CA, USA). Twelve positive clones per sample were sequenced with both forward and reverse primers using the above mentioned protocols for sequencing.

5.5.3 Genome walking protocols

Genomic DNA of *Tragopogon dubius* (diploid parental species) was digested with three different restriction enzymes: *EcoRV*, *DraI* and *ScaI* (New England Biolabs) in separate reaction tubes containing 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 reactions were ethanol precipitated in the presence of 20 µg glycogen and 3M sodium acetate. Adapter ligation to the precipitated, digested genomic DNA was performed 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 performed in 50-µl total volume using 10 mM dNTPs, 10X PCR buffer (Takara Biotechnology, Japan), 10 µM of adapter primer AP1 (Forward) and gene-specific primer (Reverse) (gene-specific reverse primers for all the genes S2, S3, S8, S18 and TDF-44 are listed in Table S5.1) and 1 unit of Takara Ex Taq polymerase (Takara Biotechnology, Japan). PCR profile for the primary PCR was as follows: first 7 cycles at 94°C for 25 sec, 72°C for 3 min, then 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 for the nested round were diluted 1:50 in ddH_2O. In the secondary PCR, 10 µM nested adapter primer AP2 (forward) and internal gene-specific primers (reverse) were used (Table S5.1), 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 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. At least ten positive clones per gene per individual were sequenced. The resulting sequences for each gene were aligned with previously obtained sequences of that gene in Sequencher. New methylation-specific primers were designed to amplify promoter regions from bisulfite-converted DNA. The amplified promoter regions from bisulfite converted DNA and genomic DNA of all five genes were sequenced for the *T. miscellus* polyploids, and the progenitors *T. dubius* and *T. pratensis*.

Table S5. 1 List of primers

Experiment	Region	Strand	Primer name	Primer sequence (5'-3')
BS-sequencing	TDF44-gene	Sense	Dub-sense-gDNA-F	GGTGAACAAATGTTAGTCTATGAGTACA
BS-sequencing	TDF44-gene	Sense	Dub-sense-gDNA-R	GTCCCCTTCACTTGAGTAGAAACATAAC
BS-sequencing	TDF44-gene	Sense	Dub-sense-F1	TGAGTATATTTTAAATGGTATTTTAAAGG
BS-sequencing	TDF44-gene	Sense	Dub-sense-F2	AGTAAGGTGAATAAATGTTAGTTTATGAGTA
BS-sequencing	TDF44-gene	Sense	Dub-sense-R1	CAATATCCCCTTCACTTAAATAAAAAC
BS-sequencing	TDF44-gene	Sense	Dub-sense-R2	AACATAACCCTTTAAATCATCTCCCAAC
BS-sequencing	TDF44-gene	Antisense	Dub-AS-gDNA-F	CAGATACTCATGTAGAGTTTACCATGGGA
BS-sequencing	TDF44-gene	Antisense	Dub-AS-gDNA-R	CAGGGGAAGTGAACTCATCTTTGTATTG
BS-sequencing	TDF44-gene	Antisense	Dub-AS-F1	TAGATATTTATGTAGAGTTTATTATGGGAT
BS-sequencing	TDF44-gene	Antisense	Dub-AS-F2	GAGTTTATTATGGGATTTTTTATTAGAAAG
BS-sequencing	TDF44-gene	Antisense	Dub-AS-R1	TTACAAAAAAAATAAACTCATCTTTAT
BS-sequencing	TDF44-gene	Antisense	Dub-AS-R2	ACTCATCTTTATATTAAAAAACTTAATAA
BS-sequencing	S2-gene	Sense	S2-sense-gdna-F	CCTTTGTACATTTTCATTCGGTAAAC
BS-sequencing	S2-gene	Sense	S2-sense-F1	GGTGTTAGGTTGAAGATGTTATTAAAG
BS-sequencing	S2-gene	Sense	S2-sense-F2	GAGGTAATAAATGGGATAATTTTAG
BS-sequencing	S2-gene	Sense	S2-sense-gdna-R	AACAAACCCAGACTGTGGACCT
BS-sequencing	S2-gene	Sense	S2-sense-R1	CCATCTTTCTTATCTCCACTCTAATC
BS-sequencing	S2-gene	Sense	S2-sense-R2	CCACTCTAATCTTATACACCATTTA
BS-sequencing	S2-gene	Antisense	S2-AS-gdna-F	ACAAACCCAGACTGTGGACCTG
BS-sequencing	S2-gene	Antisense	S2-AS -F1	TGGTTTTGTATATTATTTGATTGT
BS-sequencing	S2-gene	Antisense	S2-AS-F2	GATTGTTTTTTATTTGGGTTGTA
BS-sequencing	S2-gene	Antisense	S2-AS-gdna-R	CCTTTGTACATTTTCATTCGGTAAAC
BS-sequencing	S2-gene	Antisense	S2-AS-R1	TACCACCAAAACCCTAAAAACAACA
BS-sequencing	S2-gene	Antisense	S2-AS-R2	CAACCCTAAAAAAAAACAATCAAATC

BS-sequencing	S3-gene	Sense	S3-sense-gdna-F	AGACGTGGGAGTAATTACAAGT
BS-sequencing	S3-gene	Sense	S3-sense-F1	GTG ATT AGG GTT AGT TTG ATA AAG
BS-sequencing	S3-gene	Sense	S3-sense-F2	GGGGAGAAAGAAATGGTAATTTTAGTG
BS-sequencing	S3-gene	Sense	S3-sense-gdna-R	CAGGATTAGTGTGATGTCTCCAACAG
BS-sequencing	S3-gene	Sense	S3-sense-R1	CACATTTATTTACACATACCAAAC
BS-sequencing	S3-gene	Sense	S3-sense-R2	ACACATACCAAACTTAATATTACCTC
BS-sequencing	S3-gene	Antisense	S3-AS-gdna-F	CAACAGGATTAGTGTGATGTCTCC
BS-sequencing	S3-gene	Antisense	S3-AS -F1	GTATATGTTAAATTTGGTGTTG
BS-sequencing	S3-gene	Antisense	S3-AS-F2	TGGTGTTGTTTTTATGTATTTG
BS-sequencing	S3-gene	Antisense	S3-AS-gdna-R	GAATGGGGAGAAAGAAATGGTAAC
BS-sequencing	S3-gene	Antisense	S3-AS-R1	ACTCCAATAATCAAAACTAACTTAAC
BS-sequencing	S3-gene	Antisense	S3-AS-R2	CTAACTTAACAAAACAAATATACTCC
BS-sequencing	S8-gene	Sense	S8-sense-gdna-F	GACGCAATATAACAACCTTTCTTG
BS-sequencing	S8-gene	Sense	S8-sense-Unmethyl-F	GAAAATTATGAATATTATTGATGG
BS-sequencing	S8-gene	Sense	S8-sense-Methyl-F	GAAAATTACGAATATTATCGATGG
BS-sequencing	S8-gene	Sense	S8-sense-gdna-R	GCGGAAGTGTGTCCTGATAAAGC
BS-sequencing	S8-gene	Sense	S8-sense-Unmethyl-R	CCTAATAAAACATCATAAAAATTCC
BS-sequencing	S8-gene	Sense	S8-sense-Methyl-R	CCTAATAAAACATCGTAAAAATTCC
BS-sequencing	S8-gene	Antisense	S8-AS-gdna-F	GCGGAAGTGTGTCCTGATAAAGC
BS-sequencing	S8-gene	Antisense	S8-AS-Unmethyl-F	GTGGAAGTGTGTTTTGATAAAGTATTG
BS-sequencing	S8-gene	Antisense	S8-AS-methyl-F	GCGGAAGTGTGTTTTGATAAAGTATCG
BS-sequencing	S8-gene	Antisense	S8-AS-gdna-R	TGACGCAATATAACAACCTTTCTTG
BS-sequencing	S8-gene	Antisense	S8-AS-unmethyl-R	CAAATATCATCAATAAAAAATTACAAATC
BS-sequencing	S8-gene	Antisense	S8-AS-methyl-R	CGAATATCATCGATAAAAAATTACAAATC
BS-sequencing	S18-gene	Sense	S18-sense-gdna-F	TTGTGACTTCCCAATACTTGCTCT
BS-sequencing	S18-gene	Sense	S18-sense-Unmethyl-F	AGAAATTGTTGATGATGATTTAAG
BS-sequencing	S18-gene	Sense	S18-sense-	TTAATAGAAATCGTCGACGACG

			Methyl-F	
BS-sequencing	S18-gene	Sense	S18-sense-gdna-R	AAGACACCCTATTCACACCCAAC
BS-sequencing	S18-gene	Sense	S18-sense-R1	ACCCAACCCCATCAAATCTTTCAAT
BS-sequencing	S18-gene	Sense	S18-sense-R2	ACATCAAATATTCCAAAATCCA
BS-sequencing	S18-gene	Antisense	S18-AS-gdna-F	GTTGGGTGTGAATAGGGTGTCTT
BS-sequencing	S18-gene	Antisense	S18-AS-Unmethyl-F	ATGTATTAAGAATTTTAGTGATGAG
BS-sequencing	S18-gene	Antisense	S18-AS-methyl-F	ACGTATTAAGAATTTTAGCGACGAG
BS-sequencing	S18-gene	Antisense	S18-AS-F1	TAGGGTTTATTTTTATTGATATTTTAG
BS-sequencing	S18-gene	Antisense	S18-AS-gdna-R	GTGACTTCCCAATACTTGCTCTCGGC
BS-sequencing	S18-gene	Antisense	S18-AS-unmethyl-R	TCAACAAAAACCACCAACAACAAC
BS-sequencing	S18-gene	Antisense	S18-AS-methyl-R	TTCAACAAAAACCGCCGACGACG
BS-sequencing	S18-gene	Antisense	S18-AS-R1	TATAACTTCCCAATACTTACTCTC
BS-sequencing	S2-Promoter	Sense	S2-gdna-5'-F	GTGTAATGAACCTAGGACTG
BS-sequencing	S2-Promoter	Sense	S2-sense-5'L-F1	GGTTGGTTTAAAATAGTGTAATG
BS-sequencing	S2-Promoter	Sense	S2-sense-5'L-F2	GTGTAATGAATTTAGGATTG
BS-sequencing	S2-Promoter	Sense	S2-sense-5'L-R1	CAAATACCAAAATAAAATTATAC
BS-sequencing	S2-Promoter	Sense	S2-sense-5'L-R2	CCAAAATAAAATTATACAATC
BS-sequencing	S3-Promoter	Sense	S3-gdna-5'-F	GATGAGGGCGAGTTAGATACGACC
BS-sequencing	S3-Promoter	Sense	S3-sense-5'L-F1	GATTTTTTGGAGTATTTGAG
BS-sequencing	S3-Promoter	Sense	S3-sense-5'L-F2	GGAGTATTTGAGATTTTGTTTT
BS-sequencing	S3-Promoter	Sense	S3-sense-5'L-R1	CTTTACCTACAATATCTTCT
BS-sequencing	S3-Promoter	Sense	S3-sense-5'L-R2	CTTCTAAATCAACAATATCAAAC
BS-sequencing	S8-Promoter	Sense	S8-gdna-5'-F	GCTGTCATGTCTAAGCCGCATCG
BS-sequencing	S8-Promoter	Sense	S8-sense-5'L-F1	TATGTGTATATTGTTTTGGTGG

BS-sequencing	S8-Promoter	Sense	S8-sense-5'L-F2	GTGTATATTGTTTTGGTGGATTTGA
BS-sequencing	S8-Promoter	Sense	S8-sense-5'L-R1	CTTAATTACACAAAAATTCACAATC
BS-sequencing	S8-Promoter	Sense	S8-sense-5'L-R2	ACTAAACCCTCAAAACCCTACC
BS-sequencing	S8-Promoter	Sense	S18-gdna-5'-F	CCATTACCGTTAGAAACATTGGTC
BS-sequencing	S18-Promoter	Sense	S18-sense-5'L-F1	GGTTATGTTTTTTATTGTAAATTTATG
BS-sequencing	S18-Promoter	Sense	S18-sense-5'L-F2	GAAGAAAGAGAATTAAAATTAG
BS-sequencing	S18-Promoter	Sense	S18-sense-5'L-R1	ATTTCTATTAAAAAAAAACC
BS-sequencing	S18-Promoter	Sense	S18-sense-5'L-R2	TTAAAAAATCACAATAATAACAAC
BS-sequencing	S18-Promoter	Sense	TDF44-gdna-5'-F	CGATCGAATCTAATCCCAGTGG
BS-sequencing	TDF44-Promoter	Sense	TDF44-sense-5'L-F1	GTTGGTGGTTGTGTGATTGTATTA
BS-sequencing	TDF44-Promoter	Sense	TDF44-sense-5'L-F2	GTGTGATTGTATTATTAATAGG
BS-sequencing	TDF44-Promoter	Sense	TDF44-sense-5'L-R1	CTAACATTTATTCACCTTACTC
BS-sequencing	TDF44-Promoter	Sense	TDF44-sense-5'L-R2	CAAAACCCCACAAAAACTAC
Genome walking	--	Sense	AP1	GTAATTCGCATCACTATAGCTC
Genome walking	--	Sense	AP2	ACTATAGCTCACCGCTGGT
Genome walking	--	Sense	NA44	GTAATTCGCATCACTATAGCTCACCGCTGGTC GACGGCCCGGGCTGGT
Genome walking	--	Sense	NA45	PO4-ACCAGCCC-NH2
Genome walking	S2-5'race	Sense	S2-5'race-dub1	CAATCTTGGTTCCAGAAAAGTTAC
Genome walking	S2-5'race	Sense	S2-5'race-dub2	CTTTGGTGGCATCTTCAACCTGAC
Genome walking	S3-5'race	Sense	S3-5'race-dub1	TCACTGGAGTTACCATTTCTTTCTC
Genome walking	S3-5'race	Sense	S3-5'race-dub2	GGATTTAGGAGTATACCTGCTTTGTC
Genome walking	S8-5'race	Sense	S8-5'race-dub1	GTAATTTTCCATCGATGATATTCG
Genome walking	S8-5'race	Sense	S8-5'race-dub2	GACATGAATGTGCCACCCATCAAG

Genome walking	S18-5’race	Sense	S18-5'race-R1	CGGTTTCGAGAGGTGGGTGTTGCTC
Genome walking	S18-5’race	Sense	S18-5'race-R2	CTTCAATTTCGCGAATAAGGAGGTG
Genome walking	TDF44-5’race	Sense	TDF-44-5'race-R1	GGTGATGCCCTAAACCCTAAAC
Genome walking	TDF44-5’race	Sense	TDF-44-5'race-R2	CCTTTAGGGTACCATTTGAGATG

CHAPTER SIX

6. General discussion

The thesis comprises a detailed study mainly on genetic and some epigenetic consequences of polyploidy. In general, the evolution of polyploid genomes is associated with several genetic modifications including homeolog losses, changes in the expression pattern of duplicated genes at the transcriptome and proteome level, changes in genome size and chromosomal re-organization (Kashkush *et al.*, 2002a; Seoighe & Gehring, 2004; Tate *et al.*, 2006b; Adams, 2007; Chen, 2007a; Gaeta *et al.*, 2007; Leitch *et al.*, 2008; Buggs *et al.*, 2010a; Flagel & Wendel, 2010; Chester *et al.*, 2012; Koh *et al.*, 2012).

To follow the genetic consequences of polyploidy, our first research project was based on an analysis of cytonuclear coordination using the *rbcS/rbcL* subunit system in synthetic and natural *Tragopogon* polyploids. This is the first study of its kind on the cytonuclear coordination in naturally formed and relatively young polyploids (less than 100 years old) and, most importantly, *Tragopogon* is a system with reciprocal polyploids (short- and long-liguled *T. miscellus*) to study maternal influences on cytonuclear interaction. Homeolog losses and biases in their expression pattern, which are commonly observed in polyploid species (Bottley *et al.*, 2006; Tate *et al.*, 2006b; Flagel & Wendel, 2010; Koh *et al.*, 2010; Xiong *et al.*, 2011; Feldman *et al.*, 2012; Wendel *et al.*, 2012), were also found for the nuclear-encoded subunit rbcS. Our results were different from previously observed homeolog losses and changes in the expression patterns in *Tragopogon* as previous reports showed species-specific retention and expression of homeologs with *T. dubius* homeologs lost and silenced more frequently compared to *T. pratensis* homeologs in the reciprocally formed *T. miscellus* polyploids (Tate *et al.*, 2006b; Buggs *et al.*, 2009; Tate *et al.*, 2009a; Malinska *et al.*, 2010). Those studies did not show any maternal bias in the pattern of loss and silencing. The results from *rbcS*/*rbcL* study were in line with our hypothesis of a maternal influence on the cytonuclear coordination in polyploids. Preferential loss and silencing of the paternal homeolog of *rbcS* was observed in both short- and long-liguled natural *T. miscellus* polyploids, which retained and

expressed only the maternal homeolog of *rbcS* to potentially facilitate cytonuclear coordination with the maternally inherited chloroplast subunit *rbcL*. Demonstration of this pattern by all the independently formed populations of *T. miscellus* shows repeatability of cytonuclear coordination. Polymorphisms between the parental homeologs might be responsible for the differential selection and expression of *rbcS* homeologs. Synthetic polyploids on the other hand showed retention and expression of both parental copies of *rbcS* suggesting cytonuclear coordination do not establish immediately upon polyploid formation.

Our second study revealed that considerable genetic variation has been incorporated into synthetic polyploids by their parents. Synthetic polyploids inherited the parental loci and showed the presence of multilocus genotypes reported previously for natural polyploids. Interestingly, those synthetic genotypes which deviated mostly from those multilocus Types appeared to be less successful at or beyond F_1 and S_0 generations as revealed by Tate *et al.*, (2009b) who showed that only few crosses were successful in producing synthetic polyploids after making several attempts. This suggests a potential role for some genetic rules in the formation of successful polyploids; however, much more work is required in this area.

Based on the formation of short-liguled (*T. dubius*♂ × *T. pratensis*♀) and long-liguled (*T. dubius*♀ × *T. pratensis*♂) inflorescences (Soltis & Soltis, 1989) in both naturally formed and synthetically produced *Tragopogon miscellus* reciprocal polyploids, our third study proposed the influence of maternal factors to the formation of these long- and short-liguled forms. Comparison between progenitor orthologs of floral identity genes reported in other species of Asteraceae (Yu *et al.*, 1999; Dezar *et al.*, 2003; Shchennikova *et al.*, 2004; Shul'ga *et al.*, 2008) showed the presence of few polymorphisms that were not informative to differentiate between the parental homeologs in the polyploids. *T. dubius* has generally been reported to be more variable compared to *T. pratensis* (Soltis *et al.*, 1995a; Symonds *et al.*, 2010) showed almost complete homozygosity at those

polymorphic sites but *T. pratensis* was always heterozygous at those loci making any inference about progenitor specific expression difficult. Nonetheless, this study reports first the orthologs of the floral ABC MADS-box transcription factors for *Tragopogon* and suggests that these loci are not responsible for the differing floral and inflorescence morphologies of the reciprocally formed *T. miscellus* polyploids.

Among the epigenetic consequences of polyploidy, DNA methylation is a commonly reported mechanism responsible for homeolog silencing (Shaked *et al.*, 2001; Madlung *et al.*, 2002; Salmon *et al.*, 2008; Salmon & Ainouche, 2010; Zhang *et al.*, 2013). An examination of five homeologous loci previously reported to be silenced in *T. miscellus* polyploids (Tate *et al.*, 2006b; Buggs *et al.*, 2009) demonstrated that DNA methylation is also playing a role in the evolution *Tragopogon* polyploids. DNA methylation of *T. dubius* homeolog of TDF-44 and S2 is responsible for the silencing of these loci as reported in Tate *et al.* (2006b) and Buggs *et al.* (2009), respectively. All the reciprocally formed *T. miscellus* individuals studied showed this silencing pattern, suggesting that there was not a maternal influence involved in this case. These results are also similar to previous studies in *Tragopogon* which showed frequent loss and silencing of *T. dubius* homeologs (Kovarik *et al.*, 2005; Tate *et al.*, 2006b; Matyasek *et al.*, 2007; Buggs *et al.*, 2009; Buggs *et al.*, 2010a; Buggs *et al.*, 2012a). The remaining three homeologous loci (S3, S8 and S18) examined did not show DNA methylation suggesting other mechanisms like histone deacetylation (causing transcriptional repression) and RNA interference (involved in post-transcriptional silencing) may be responsible for their silencing (Carthew & Sontheimer, 2009; Kim *et al.*, 2012; Feng & Guang, 2013; Ma *et al.*, 2013).

6.1 Conclusions and future perspectives

The studies described here suggest that polyploidy has led to considerable genetic and epigenetic modifications in the natural polyploids of *Tragopogon*. Natural polyploids from various independently formed natural populations have shown the same pattern of

genetic and epigenetic alterations proposing a patterned response of natural polyploids from different origins. The maternal influence on the cytonuclear coordination observed here, in terms of preferential retention and expression of maternal homeolog of *rbcS,* suggests regulatory coordination is required to maintain genomic balance among nuclear and cytoplasmic counterparts. Further studies based on protein-protein interaction among *rbcS/rbcL* in *T. miscellus* polyploids and analysis of other cytonuclear complexes would improve our understanding of this trend in cytonuclear coordination. The overall trend in synthetic (S_1) polyploids was towards additivity in retention and expression of both progenitor homeologs as shown by the above studies and a previous study by Buggs *et al.* (2009). The absence of such genetic changes in the synthetic polyploids of *Tragopogon* suggests that those changes take time to establish. Examination of subsequent generations of synthetic polyploids would help us to know the timing of change.

Surprisingly, the transcriptomic study did not reveal any significant differences between the long- and short-liguled forms of *T. miscellus* polyploids in terms of differential expression patterns of progenitor copies for the orthologs of floral identity genes. Since only a few floral identity genes were investigated in this preliminary study, a complete analysis of the resulting transcriptomic data generated from the inflorescences of the synthetic and natural short- and long-liguled *T. miscellus* polyploids would yield valuable findings. Moreover, a detailed developmental study of the individual floral whorls and inflorescences (using morphological and transcriptomic analysis) would be warranted to identify the genes responsible for these differing forms.

The epigenetic study involving an analysis of DNA methylation status of *T. miscellus* polyploids showed its contribution to the evolution of *Tragopogon* polyploids. Although we found silencing of only two homeologous loci by DNA methylation out of five loci studied, employing genome-wide high throughput analysis of DNA methylation would clarify its role in *Tragopogon* evolution.

Bibliography

Adams K, Cronn R, Percifield R, Wendel J. 2003a. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proc Natl Acad Sci USA* **100**: 4649 - 4654.

Adams K, Wendel J. 2005a. Polyploidy and genome evolution in plants. *Curr Opin Plant Biol* **8**: 135 - 141.

Adams KL. 2007. Evolution of duplicate gene expression in polyploid and hybrid plants. *Journal of Heredity* **98**: 136-141.

Adams KL, Cronn R, Percifield R, Wendel JF. 2003b. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proceedings of the National Academy of Sciences of the United States of America* **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. 2004. Exploring the genomic mysteries of polyploidy in cotton. *Biological Journal of the Linnean Society* **82**: 573-581.

Adams KL, Wendel JF. 2005b. Novel patterns of gene expression in polyploid plants. *Trends in Genetics* **21**: 539-543.

Akhunov ED, Sehgal S, Liang H, Wang S, Akhunova AR, Kaur G, Li W, Forrest KL, See D, Simkova H, *et al.* 2013. Comparative analysis of syntenic genes in grass genomes reveals accelerated rates of gene structure and coding sequence evolution in polyploid Wheat. *Plant Physiology* **161**: 252-265.

Albert VA, Soltis DE, Carlson JE, Farmerie WG, Wall PK, Ilut DC, Solow TM, Mueller LA, Landherr LL, Hu Y, *et al.* 2005. Floral gene resources from basal angiosperms for comparative genomics research. *BMC Plant Biology* **5**.

Angenent GC, Busscher M, Franken J, Mol JNM, Vantunen AJ. 1992. Differential expression of 2 MADS box genes in wild-type and mutant *Petunia* flowers. *Plant Cell* **4**: 983-993.

Anssour S, Baldwin IT. 2010. Variation in antiherbivore defense responses in synthetic *Nicotiana* allopolyploids correlates with changes in uniparental patterns of gene expression. *Plant Physiology* **153**: 1907-1918.

Anssour S, Krugel T, Sharbel TF, Saluz HP, Bonaventure G, Baldwin IT. 2009. Phenotypic, genetic and genomic consequences of natural and synthetic polyploidization of *Nicotiana attenuata* and *Nicotiana obtusifolia*. *Annals of Botany* **103**: 1207-1217.

APGIII. 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Botanical Journal of the Linnean Society* **141**: 399-436.

Aronesty E. 2013. TOBioiJ : Comparison of Sequencing Utility Programs. *The Open Bioinformatics Journal* **7**: 1-8.

Ashton PA, Abbott RJ. 1992. Multiple origins and genetic diversity in the newly arisen allopolyploid species, *Senecio Cambrensis Rosser* (Compositae). *Heredity* **68**: 25-32.

Balao F, Herrera J, Talavera S. 2011. Phenotypic consequences of polyploidy and genome size at the microevolutionary scale: a multivariate morphological approach. *New Phytologist* **192**: 256-265.

Barcaccia G, Tavoletti S, Mariani A, Veronesi F. 2003. Occurrence, inheritance and use of reproductive mutants in alfalfa improvement. *Euphytica* **133**: 37-56.

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

Barr CM, Fishman L. 2011. The nuclear component of a cytonuclear hybrid incompatibility in *Mimulus* maps to a cluster of pentatricopeptide repeat genes. *Genetics* **184**: 455-465.

Baumel A, Ainouche ML, Levasseur JE. 2001. Molecular investigations in populations of *Spartina anglica C.E. Hubbard* (Poaceae) invading coastal Brittany (France). *Molecular Ecology* **10**: 1689-1701.

Beaulieu J, Jean M, Belzile F. 2009. The allotetraploid *Arabidopsis thaliana-Arabidopsis lyrata* subsp *petraea* as an alternative model system for the study of polyploidy in plants. *Molecular Genetics and Genomics* **281**: 421-435.

Benavente E, Orellana J. 1991. Chromosome differentiation and pairing behavior of polyploids: an assessment on preferential metaphase I associations in colchicine-induced autotetraploid hybrids within the genus *Secale*. *Genetics* **128**: 433-442.

Bhalla N, Dernburg AF. 2008. Prelude to a division. *Annu Rev Cell Dev Biol* **24**: 397-424.

Birchler JA, Veitia RA. 2010. The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. *New Phytologist* **186**: 54-62.

Birky CW. 1995. Uniparental inheritance of mitochondrial and chloroplast genes - mechanisms and evolution. *Proceedings of the National Academy of Sciences of the United States of America* **92**: 11331-11338.

Blackman BK, Rasmussen DA, Strasburg JL, Raduski AR, Burke JM, Knapp SJ, Michaels SD, Rieseberg LH. 2011. Contributions of flowering time genes to sunflower domestication and improvement. *Genetics* **187**: 271-287.

Blanc G, Wolfe K. 2004a. Widespread paleopolyploidy in model plant species inferred from age distribution of duplicated genes. *Plant Cell* **16**: 1667 - 1678.

Blanc G, Wolfe KH. 2004b. Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *Plant Cell* **16**: 1679-1691.

Bombarely A, Edwards KD, Sanchez-Tamburrino J, Mueller LA. 2012. Deciphering the complex leaf transcriptome of the allotetraploid species *Nicotiana tabacum*: a phylogenomic perspective. *BMC Genomics* **13**: (17 August 2012)-(2017 August 2012).

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

Bowman JL, Smyth DR, Meyerowitz EM. 1989. Genes directing flower development in *Arabidopsis*. *The Plant Cell Online* **1**: 37-52.

Bowman JL, Smyth DR, Meyerowitz EM. 1991. Genetic interactions among floral homeotic genes of *Arabidopsis*. *Development* **112**: 1-20.

Braasch I, Postlethwait JH 2012. Polyploidy in fish and the teleost genome duplication. In: D. E. SoltisP. S. Soltis eds. *Polyploidy and genome evolution*. Berlin Heidelberg: Springer, 341-383.

Bradley D, Carpenter R, Sommer H, Hartley N, Coen E. 1993. Complementary floral homeotic phenotypes result from opposite orientations of a transposon at the *PLENA* locus of *Antirrhinum*. *Cell* **72**: 85-95.

Brandon SG. 2002. Tansley Review No. 132. Evolutionary Dynamics of Grass Genomes. *New Phytologist* **154**: 15-28.

Bremer K 1994. Asteraceae: Cladistics and classification. Portland: Timber press.

Bretagnolle F, Thompson JD. 1995. Tansley review no. 78. Gametes with the somatic chromosome number: mechanisms of their formation and role in the evolution of atotpolyploid plants. *New Phytololgist* **129**: 1-22.

Broholm SK, Tahtiharju S, Laitinen RA, Albert VA, Teeri TH, Elomaa P. 2008. A TCP domain transcription factor controls flower type specification along the radial axis of the *Gerbera* (Asteraceae) inflorescence. *Proc Natl Acad Sci U S A* **105**: 9117-9122.

Brownfield L, Kohler C. 2011. Unreduced gamete formation in plants: mechanisms and prospects. *Journal of Experimental Botany* **62**: 1659-1668.

Buggs R, Elliott N, Zhang L, Koh J, Viccini L, Soltis D, Soltis P. 2010a. Tissue-specific silencing of homoeologs in natural populations of the recent allopolyploid Tragopogon mirus. *New Phytol* **186**: 175 - 183.

Buggs RJA. 2008. Towards natural polyploid model organisms. *Molecular Ecology* **17**: 1875-1876.

Buggs RJA, Chamala S, Wu W, Gao L, May GD, Schnable PS, Soltis DE, Soltis PS, Barbazuk WB. 2010b. Characterization of duplicate gene evolution in the recent natural allopolyploid *Tragopogon miscellus* by next-generation sequencing and Sequenom iPLEX MassARRAY genotyping. *Molecular Ecology* **19**: 132-146.

Buggs RJA, Chamala S, Wu W, Tate Jennifer A, Schnable Patrick S, Soltis Douglas E, Soltis PS, Barbazuk WB. 2012a. Rapid, repeated, and clustered loss of duplicate genes in allopolyploid plant populations of independent origin. *Current biology* **22**: 248-252.

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

Buggs RJA, Renny-Byfield S, Chester M, Jordon-Thaden IE, Viccini LF, Chamala S, Leitch AR, Schnable PS, Barbazuk WB, Soltis PS, *et al.* 2012b. Next-generation sequencing and genome evolution in allopolyploids. *American Journal of Botany* **99**: 372-382.

Buggs RJA, Zhang LJ, Miles N, Tate JA, Gao L, Wei W, Schnable PS, Barbazuk WB, Soltis PS, Soltis DE. 2011. Transcriptomic shock generates evolutionary novelty in a newly formed, natural allopolyploid plant. *Current biology* **21**: 551-556.

Burge C, Karlin S. 1997. Prediction of complete gene structures in human genomic DNA. *Journal of Molecular Biology* **268**: 78-94.

Cam HP. 2010. Roles of RNAi in chromatin regulation and epigenetic inheritance. *Epigenomics* **2**: 613-626.

Cañestro C 2012. Two rounds of whole-genome duplication: evidence and impact on the evolution of vertebrate innovations. In: D. E. SoltisP. S. Soltis eds. *Polyploidy and genome evolution*. Berlin Heidelberg: Springer, 309-339.

Carputo D, Barone A, Frusciante L. 2000. 2n gametes in the potato: essential ingredients for breeding and germplasm transfer. *Theoretical and Applied Genetics* **101**: 805 - 813.

Carthew RW, Sontheimer EJ. 2009. Origins and mechanisms of miRNAs and siRNAs. *Cell* **136**: 642-655.

Castro M, Castro S, Loureiro J. 2012. Genome size variation and incidence of polyploidy in *Scrophulariaceae sensu lato* from the Iberian Peninsula. *AoB plants* **2012**: pls037-pls037.

Chae WB, Hong SJ, Gifford JM, Rayburn AL, Widholm JM, Juvik JA. 2012. Synthetic polyploid production of *Miscanthus sacchariflorus*, *Miscanthus sinensis*, and *Miscanthus* x *giganteus*. *Global Change Biology Bioenergy* **5**: 338-350.

Chague V, Just J, Mestiri I, Balzergue S, Tanguy A, Huneau C, Huteau V, Belcram H, Coriton O, Jahier J, *et al.* 2010. Genome-wide gene expression changes in genetically stable synthetic and natural wheat allohexaploids. *New Phytol* **187**: 1181 - 1194.

Chan SWL, Henderson IR, Jacobsen SE. 2005. Gardening the genome: DNA methylation in *Arabidopsis thaliana*. *Nature Reviews Genetics* **6**: 351-360.

Chang WC, Lee TY, Huang HD, Huang HY, Pan RL. 2008. PlantPAN: Plant promoter analysis navigator, for identifying combinatorial cis-regulatory elements with distance constraint in plant gene groups. *BMC Genomics* **9**: 561.

Chantret N, Salse J, Sabot F, Rahman S, Bellec A, Laubin B, Dubois I, Dossat C, Sourdille P, Joudrier P, *et al.* 2005. Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (Triticum and aegilops). *Plant Cell* **17**: 1033-1045.

Chapman BA, Bowers JE, Feltus FA, Paterson AH. 2006. Buffering of crucial functions by paleologous duplicated genes may contribute cyclicality to angiosperm genome duplication. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 2730-2735.

Chapman MA, Leebens-Mack JH, Burke JM. 2008. Positive selection and expression divergence following gene duplication in the sunflower *CYCLOIDEA* gene family. *Molec Biol Evol* **25**: 1260-1273.

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

Chen F, Beecher BS, Morris CF. 2010. Physical mapping and a new variant of *Puroindoline B-2* genes in wheat. *Theoretical and Applied Genetics* **120**: 745-751.

Chen Z. 2007a. Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annu Rev Plant Biol* **58**: 377 - 406.

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

Chen ZJ, Pikaard CS. 1997. Transcriptional analysis of nucleolar dominance in polyploid plants: Biased expression/silencing of progenitor rRNA genes is developmentally regulated in *Brassica*. *Proceedings of the National Academy of Sciences of the United States of America* **94**: 3442-3447.

Chen ZJ, Scheffler BE, Dennis E. 2007. Toward sequencing Cotton (*Gossypium*) genomes. *Plant Physiology* **145**: 1303-1310.

Cheng XF, Wittich PE, Kieft H, Angenent G, XuHan X, van Lammeren AAM. 2000. Temporal and spatial expression of MADS box genes, *FBP7* and *FBP11*, during initiation and early development of ovules in wild type and mutant *Petunia hybrida*. *Plant Biology* **2**: 693-702.

Chester M, Gallagher JP, Symonds VV, da Silva AVC, Mavrodiev EV, Leitch AR, Soltis PS, Soltis DE. 2011. Extensive chromosomal variation in a recently formed natural allopolyploid species, Tragopogon miscellus (Asteraceae). *Proceedings of the National Academy of Sciences of the United States of America* **109**: 1176-1181.

Chester M, Gallagher JP, Symonds VV, da Silva AVC, Mavrodiev EV, Leitch AR, Soltis PS, Soltis DE. 2012. Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon miscellus* (Asteraceae). *Proceedings of the National Academy of Sciences of the United States of America* **109**: 1176-1181.

Chester M, Lipman MJ, Gallagher JP, Soltis PS, Soltis DE. 2013. An assessment of karyotype restructuring in the neoallotetraploid *Tragopogon miscellus* (Asteraceae). *Chromosome Research* **21**: 75-85.

Clark SJ, Harrison J, Paul CL, Frommer M. 1994. High-sensitivity mapping of methylated cytosines. *Nucleic Acids Research* **22**: 2990-2997.

Clarkson JJ, Knapp S, Garcia VF, Olmstead RG, Leitch AR, Chase MW. 2004. Phylogenetic relationships in Nicotiana (Solanaceae) inferred from multiple plastid DNA regions. *Molecular Phylogenetics and Evolution* **33**: 75-90.

Clegg MT. 1993. Chloroplast gene-sequences and the study of plant evolution. *Proceedings of the National Academy of Sciences of the United States of America* **90**: 363-367.

Coen ES, Meyerowitz EM. 1991. The war of the whorls - genetic interactions controlling flower development. *Nature* **353**: 31-37.

Coen ES, Nugent JM, Luo D, Bradley D, Cubas P, Chadwick M, Copsey L, Carpenter R. 1995. Evolution of floral symmetry. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **350**: 35-38.

Coen ES, Romero JM, Doyle S, Elliott R, Murphy G, Carpenter R. 1990. *FLORICAULA*: a homeotic gene required for flower development in *Antirrhinum majus*. *Cell* **63**: 1311-1322.

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

Comai L. 2005b. The advantages and disadvantages of being polyploid. *Nat Rev Genet* **6**: 836 - 846.

Comai L, Tyagi AP, Winter K, Holmes-Davis R, Reynolds SH, Stevens Y, Byers B. 2000. Phenotypic instability and rapid gene silencing in newly formed *Arabidopsis* allotetraploids. *Plant Cell* **12**: 1551-1567.

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

Correns C 1909. Vererbungsversuche mit blass (gelb) grünen und buntblättrigen Sippen bei *Mirabilis, Urtica* und *Lunaria*. 291–329.

Cox MP, Peterson DA, Biggs PJ. 2010. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics* **11**: 1471-2105.

Cubas P. 2004. Floral zygomorphy, the recurring evolution of a successful trait. *BioEssays* **26**: 1175-1184.

D'Amato F. 1952. Polyploidy in the differentiation and function of tissues and cells in plants. *Caryologia* **4**: 311–358.

Davies B, Motte P, Keck E, Saedler H, Sommer H, Schwarz-Sommer Z. 1999. *PLENA* and *FARINELLI*: redundancy and regulatory interactions between two *Antirrhinum* MADS-box factors controlling flower development. *Embo Journal* **18**: 4023-4034.

De Smet R, Van de Peer Y. 2012. Redundancy and rewiring of genetic networks following genome-wide duplication events. *Current Opinion in Plant Biology* **15**: 168-176.

De Storme N, Geelen D. 2013. Sexual polyploidization in plants – cytological mechanisms and molecular regulation. *New Phytologist* **198**: 670-684.

Dean C, Pichersky E, Dunsmuir P. 1989. Structure, evolution and regulation of *rbcS* genes in higher plants. *Annual Review of Plant Physiology and Plant Molecular Biology* **40**: 415-439.

Dezar CA, Tioni MF, Gonzalez DH, Chan RL. 2003. Identification of three MADS-box genes expressed in sunflower capitulum. *Journal of Experimental Botany* **54**: 1637-1639.

Dieringer D, Schlotterer C. 2003. Microsatellite Analyzer (MSA): A platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes* **3**: 167-169.

Domingues DS, Cruz GMQ, Metcalfe CJ, Nogueira FTS, Vicentini R, Alves CdS, Van Sluys M-A. 2012. Analysis of plant LTR-retrotransposons at the fine-scale family level reveals individual molecular patterns. *BMC Genomics* **13**.

Dong S, Adams KL. 2011a. Differential contributions to the transcriptome of duplicated genes in response to abiotic stresses in natural and synthetic polyploids. *New Phytologist* **190**: 1045-1057.

Dong SW, Adams KL. 2011b. Differential contributions to the transcriptome of duplicated genes in response to abiotic stresses in natural and synthetic polyploids. *New Phytologist* **190**: 1045-1057.

Doyle GG. 1963. Preferential pairing in structural heterozygotes of *zea mays*. *Genetics* **48**: 1011-1027.

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

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

Duchoslav M, Safarova L, Jandova M. 2013. Role of adaptive and non-adaptive mechanisms forming complex patterns of genome size variation in six cytotypes of polyploid *Allium oleraceum* (Amaryllidaceae) on a continental scale. *Annals of Botany* **111**: 419-431.

Edger PP, Pires JC. 2009. Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. *Chromosome Research* **17**: 699-717.

Edmands S, Burton RS. 1999. Cytochrome C oxidase activity in interpopulation hybrids of a marine copepod: A test for nuclear-nuclear or nuclear-cytoplasmic coadaptation. *Evolution* **53**: 1972-1978.

Endress PK. 1998. *Antirrhinum* and Asteridae--evolutionary changes of floral symmetry. *Symp Soc Exp Biol* **51**: 133-140.

Fawcett J, Maere S, Van de Peer Y. 2009. Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. *Proc Natl Acad Sci USA* **106**: 5737 - 5742.

Feldman M, Levy A, Chalhoub B, Kashkush K 2012. Genomic plasticity in polyploid wheat. In: P. S. SoltisD. E. Soltis eds. *Polyploidy and genome evolution*. Berlin Heidelberg: Springer 109-135.

Feldman M, Liu B, Segal G, Abbo S, Levy AA, Vega JM. 1997. Rapid elimination of Low-Copy DNA sequences in polyploid Wheat: A possible mechanism for differentiation of homoeologous chromosomes. *Genetics* **147**: 1381-1387

Feng X, Guang S. 2013. Small RNAs, RNAi and the inheritance of gene silencing in *Caenorhabditis elegans*. *Journal of Genetics and Genomics* **40**: 153-160.

Ferrario S, Immink RGH, Shchennikova A, Busscher-Lange J, Angenent GC. 2003. The MADS box gene *FBP2* is required for *SEPALLATA* function in *Petunia*. *Plant Cell* **15**: 914-925.

Finnegan EJ, Genger RK, Peacock WJ, Dennis ES. 1998. DNA methylation in plants. *Annual Review of Plant Physiology and Plant Molecular Biology* **49**: 223-247.

Fishman L, Willis JH. 2006. A cytonuclear incompatibility causes anther sterility in *Mimulus* hybrids. *Evolution* **60**: 1372-1381.

Flagel LE, Wendel JF. 2010. Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation. *New Phytologist* **186**: 184-193.

Fukuda M, Matsuo S, Kikuchi K, Kawazu Y, Fujiyama R, Honda I. 2011. Isolation and functional characterization of the *FLOWERING LOCUS T* homolog, the *LsFT* gene, in lettuce. *J Plant Physiol* **168**: 1602-1607.

Gaeta R, Pires J. 2010. Homoeologous recombination in allopolyploids: the polyploid ratchet. *New Phytologist* **186**: 18 - 28.

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.

Genereux DP, Johnson WC, Burden AF, Stoger R, Laird CD. 2008. Errors in the bisulfite conversion of DNA: modulating inappropriate- and failed-conversion frequencies. *Nucleic Acids Res* **37**: e150.

Giuliano G, Pichersky E, Malik VS, Timko MP, Scolnik PA, Cashmore AR. 1988. An evolutionarily conserved protein-binding sequence upstream of a plant light-regulated gene. *Proceedings of the National Academy of Sciences of the United States of America* **85**: 7089-7093.

Gong L, Salmon A, Yoo M-J, Grupp KK, Wang Z, Paterson AH, Wendel JF. 2012. The cytonuclear dimension of allopolyploid evolution: An example from cotton using rubisco. *Molecular Biology and Evolution* **29**: 3023-3036.

Goto K, Meyerowitz EM. 1994. Function and regulation of the *Arabidopsis* floral homeotic gene *PISTILLATA*. *Genes & Development* **8**: 1548-1560.

Goumenaki E, Taybi T, Borland A, Barnes J. 2010. Mechanisms underlying the impacts of ozone on photosynthetic performance. *Environmental and Experimental Botany* **69**: 259-266.

Grant V. 1956. The genetic structure of races and species in *Gilia*. *Advances in Genetics* **8**: 55-87.

Grover CE, Gallagher JP, Szadkowski EP, Yoo MJ, Flagel LE, Wendel JF. 2012. Homoeolog expression bias and expression level dominance in allopolyploids. *New Phytologist* **196**: 966-971.

Grunau C, Clark SJ, Rosenthal A. 2001. Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. *Nucleic Acids Research* **29**: e65-65.

Guang S, Bochner AF, Burkhart KB, Burton N, Pavelec DM, Kennedy S. 2010. Small regulatory RNAs inhibit RNA polymerase II during the elongation phase of transcription. *Nature* **465**: 1097-1101.

Gygi SP, Rochon Y, Franza BR, Aebersold R. 1999. Correlation between protein and mRNA abundance in yeast. *Molecular and Cellular Biology* **19**: 1720-1730.

Ha M, Lu J, Tian L, Ramachandran V, Kasschau KD, Chapman EJ, Carrington JC, Chen X, Wang X-J, Chen ZJ. 2009. Small RNAs serve as a genetic buffer against genomic shock in Arabidopsis interspecific hybrids and allopolyploids.

Proceedings of the National Academy of Sciences of the United States of America **106**: 17835-17840.

Halfmann R, Lindquist S. 2010. Epigenetics in the extreme: prions and the inheritance of environmentally acquired traits. *Science* **330**: 629-632.

Hannweg K, Penter M, Sippel A. 2012. Use of polyploidy in tropical and subtropical plant improvement programmes. *Acta Hort. (ISHS)* **935**: 67-73.

Harlan JR, Dewet JMJ. 1975. On O. Winge and a prayer: origins of polyploidy. *Botanical Review* **41**: 361-390.

Harris E. 1995. Inflorescence and floral ontogeny in asteraceae: A synthesis of historical and current concepts. *The Botanical Review* **61**: 93-278.

Hauber DP, Reeves A, Stack SM. 1999. Synapsis in a natural autotetraploid. *Genome* **42**: 936-949.

He G, Zhu X, Elling AA, Chen L, Wang X, Guo L, Liang M, He H, Zhang H, Chen F, *et al.* 2010. Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids. *The Plant Cell Online* **22**: 17-33.

He X-J, Chen T, Zhu J-K. 2011. Regulation and function of DNA methylation in plants and animals. *Cell Research* **21**: 442-465.

Hegarty M, Barker G, Wilson I, Abbott R, Edwards K, Hiscock S. 2006a. Transcriptome shock after interspecific hybridization in Senecio is ameliorated by genome duplication. *Current biology* **16**: 1652 - 1659.

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

Hegarty MJ, Hiscock SJ. 2008. Genomic clues to the evolutionary success of polyploid plants. *Current biology* **18**: R435-R444.

Hegarty MJ, Jones JM, Wilson ID, Barker GL, Coghill JA, Sanchez-Baracaldo P, Liu GQ, Buggs RJA, Abbott RJ, Edwards KJ, *et al.* 2005. Development of anonymous cDNA microarrays to study changes to the *Senecio* floral transcriptome during hybrid speciation. *Molecular Ecology* **14**: 2493-2510.

Hiesey WM, M. A. Nobs, Bj¨orkman O 1971. Biosystematics, genetics, and physiological ecology of the *Erythranthe* section of *Mimulus*.In. *Experimental studies on the nature of species.*: Carnegie Inst. Wash. Publ. 1–213.

Higgins J, Magusin A, Trick M, Fraser F, Bancroft I. 2012. Use of mRNA-seq to discriminate contributions to the transcriptome from the constituent genomes of the polyploid crop species *Brassica napus*. *BMC Genomics* **13**.

Higo K, Ugawa Y, Iwamoto M, Korenaga T. 1999. Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res* **27**: 297-300.

Hileman LC, Cubas P. 2009. An expanded evolutionary role for flower symmetry genes. *Journal of Biology* **8**: 91-94.

Hillis DM, Moritz C, Mable BK, eds. 1996. *Molecular systematics*. Sunderland, MA, USA: Sinauer Associates, Inc.

Hovav R, Udall JA, Chaudhary B, Rapp R, Flagel L, Wendell JF. 2008. Partitioned expression of duplicated genes during development and evolution of a single cell in a polyploid plant. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 6191-6195.

Hu Z, Han Z, Song N, Chai L, Yao Y, Peng H, Ni Z, Sun Q. 2013. Epigenetic modification contributes to the expression divergence of three TaEXPA1 homoeologs in hexaploid wheat (*Triticum aestivum*). *New Phytologist* **197**: 1344-1352.

Hudson CM, Conant GC 2012. Yeast as a window into changes in genome complexity due to polyploidization. In: D. E. SoltisP. S. Soltis eds. *Polyploidy and genome evolution*. Berlin Heidelberg: Springer, 293-308.

Huijser P, Klein J, Lonnig WE, Meijer H, Saedler H, Sommer H. 1992. Bracteomania, an inflorescence anomaly, is caused by the loss of function of the MADS-box gene *SQUAMOSA* in *Antirrhinum majus*. *Embo Journal* **11**: 1239-1249.

Huijser P, Schmid M. 2011. The control of developmental phase transitions in plants. *Development* **138**: 4117-4129.

Hunt AG. 1994. Messenger-RNA 3' end formation in plants. *Annual Review of Plant Physiology and Plant Molecular Biology* **45**: 47-60.

Hunt AG, Macdonald MH. 1989. Deletion analysis of the polyadenylation signal of a pea ribulose-1,5-bisphosphate carboxylase small-subunit gene. *Plant Molecular Biology* **13**: 125-138.

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**: 254-267.

Izumi M, Tsunoda H, Suzuki Y, Makino A, Ishida H. 2012. *RBCS1A* and *RBCS3B*, two major members within the *Arabidopsis RbcS* multigene family, function to yield sufficient rubisco content for leaf photosynthetic capacity. *Journal of Experimental Botany* **63**: 2159-2170.

Jack T, Fox GL, Meyerowitz EM. 1994. *Arabidopsis* homeotic gene *APETALA3* ectopic expression - transcriptional and posttranscriptional regulation determine floral organ identity. *Cell* **76**: 703-716.

Jackson S, Chen ZJ. 2010. Genomic and expression plasticity of polyploidy. *Current Opinion in Plant Biology* **13**: 153-159.

Jakob SS, Meister A, Blattner FR. 2004. The considerable genome size variation of *Hordeum* Species (Poaceae) is linked to phylogeny, life form, ecology, and speciation rates. *Molecular Biology and Evolution* **21**: 860-869.

Ji L, Chen X. 2012. Regulation of small RNA stability: methylation and beyond. *Cell Research* **22**: 624-636.

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, *et al.* 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* **473**: 97-100.

Jofuku KD, den Boer BG, Van Montagu M, Okamuro JK. 1994. Control of *Arabidopsis* flower and seed development by the homeotic gene *APETALA2*. *The Plant Cell Online* **6**: 1211-1225.

Julie AL, Steven EJ. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics* **11**: 204-220.

Kadir ZBA. 1976. DNA evolution in the genus *Gossypium*. *Chromosoma* **56**: 85-94.

Kang HG, Jeon JS, Lee S, An GH. 1998. Identification of class B and class C floral organ identity genes from rice plants. *Plant Molecular Biology* **38**: 1021-1029.

Kapralov MV, Kubien DS, Andersson I, Filatov DA. 2011. Changes in rubisco kinetics during the evolution of C4 photosynthesis in *Flaveria* (Asteraceae) are associated with positive selection on genes encoding the enzyme. *Molecular Biology and Evolution* **28**: 1491-1503.

Kashkush K, Feldman M, Levy A. 2002a. Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. *Genetics* **160**: 1651 - 1659.

Kashkush K, Feldman M, Levy AA. 2002b. Gene loss, silencing and activation in a newly synthesized Wheat allotetraploid. *Genetics* **160**: 1651–1659

Kashkush K, Feldman M, Levy AA. 2003. Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nature Genetics* **33**: 102-106.

Keck E, McSteen P, Carpenter R, Coen E. 2003. Separation of genetic functions controlling organ identity in flowers. *Embo J* **22**: 1058-1066.

Kelley LA, Sternberg MJE. 2009. Protein structure prediction on the web: a case study using the *Phyre* server. *Nature Protocols* **4**: 363-371.

Ketting RF. 2011. The many faces of RNAi. *Developmental Cell* **20**: 148-161.

Kidwell MG 2005. Transposable elements. In: T. R. Gregory ed. *The evolution of the plant genome*: Elsevier Academic Press, 165-221.

Kim J-M, To TK, Seki M. 2012. An epigenetic integrator: new insights into genome regulation, environmental stress responses and developmental controls by *HISTONE DEACETYLASE 6*. *Plant and Cell Physiology* **53**: 794-800.

Koch MF. 1930. Studies in the anatomy and morphology of the composite flower I. The corolla. *American Journal of Botany* **17**: 938-952.

Koh J, Chen S, Zhu N, Yu F, Soltis PS, Soltis DE. 2012. Comparative proteomics of the recently and recurrently formed natural allopolyploid *Tragopogon mirus* (Asteraceae) and its parents. *New Phytologist* **196**: 292-305.

Koh J, Soltis PS, Soltis DE. 2010. Homeolog loss and expression changes in natural populations of the recently and repeatedly formed allotetraploid *Tragopogon mirus* (Asteraceae). *BMC Genomics* **11**: 97.

Komaki MK, Okada K, Nishino E, Shimura Y. 1988. Isolation and characterization of novel mutants of *Arabidopsis thaliana* defective in flower development. *Development* **104**: 195-&.

Kong F, Mao S, Jiang J, Wang J, Fang X, Wang Y. 2011. Proteomic changes in newly synthesized *Brassica napus* allotetraploids and their early generations. *Plant Molecular Biology Reporter* **29**: 927-935.

Kovarik A, Matyasek R, Lim KY, Skalicka K, Koukalova B, Knapp S, Chase M, Leitch AR. 2004. Concerted evolution of 18-5.8-26S rDNA repeats in *Nicotiana* allotetraploids. *Biological Journal of the Linnean Society* **82**: 615-625.

Kovarik A, Pires JC, Leitch AR, Lim KY, Sherwood AM, Matyasek R, Rocca J, Soltis DE, Soltis PS. 2005. Rapid concerted evolution of nuclear ribosomal DNA in two *Tragopogon* allopolyploids of recent and recurrent origin. *Genetics* **169**: 931-944.

Kraitshtein Z, Yaakov B, Khasdan V, Kashkush K. 2010. Genetic and epigenetic dynamics of a retrotransposon after allopolyploidization of Wheat. *Genetics* **186**: 801-812.

Krebbers E, Seurinck J, Herdies L, Cashmore A, Timko M. 1988. Four genes in two diverged subfamilies encode the ribulose-1,5-bisphosphate carboxylase small subunit polypeptides of *Arabidopsis thaliana*. *Plant Molecular Biology* **11**: 745-759.

Krizek BA, Fletcher JC. 2005. Molecular mechanisms of flower development: An armchair guide. *Nature Reviews Genetics* **6**: 688-698.

Lagercrantz U. 1998. Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that *Brassica* genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. *Genetics* **150**: 1217-1228.

Lam E, Chua NH. 1989. *asf-2* - a factor that binds to the cauliflower mosaic virus-35s promoter and a conserved GATA motif in cab promoters. *Plant Cell* **1**: 1147-1156.

Lawrence RJ, Pikaard CS. 2003. Transgene-induced RNA interference: a strategy for overcoming gene redundancy in polyploids to generate loss-of-function mutations. *Plant Journal* **36**: 114-121.

Lee D, Shin C. 2012. MicroRNA–target interactions: new insights from genome-wide approaches. *Annals of the New York Academy of Sciences* **1271**: 118-128.

Leitch IJ, Bennett MD. 2004. Genome downsizing in polyploid plants. *Biological Journal of the Linnean Society* **82**: 651-663.

Leitch IJ, Hanson L, Lim KY, Kovarik A, Chase MW, Clarkson JJ, Leitch AR. 2008. The ups and downs of genome size evolution in polyploid species of *Nicotiana* (Solanaceae). *Annals of Botany* **101**: 805-814.

Levin DA. 2003. The cytoplasmic factor in plant speciation. *Systematic Botany* **28**: 5-11.

Levy AA, Feldman M. 2004. Genetic and epigenetic reprogramming of the wheat genome upon allopolyploidization. *Biological Journal of the Linnean Society* **82**: 607-613.

Li BH, Kadura I, Fu DJ, Watson DE. 2004. Genotyping with TaqMAMA. *Genomics* **83**: 311-320.

Li QS, Hunt AG. 1997. The polyadenylation of RNA in plants. *Plant Physiology* **115**: 321-325.

Li X, Hu D, Luo M, Zhu M, Li X, Luo F, Li J, Yan J. 2013. Nuclear DNA content variation of three *Miscanthus* species in China. *Genes & Genomics* **35**: 13-20.

Lim KY, Soltis DE, Soltis PS, Tate J, Matyasek R, Srubarova H, Kovarik A, Pires JC, Xiong Z, Leitch AR. 2008. Rapid chromosome evolution in recently formed polyploids in *Tragopogon* (Asteraceae). *Plos One* **3**: e3353.

Liu B, Davis TM. 2011. Conservation and loss of ribosomal RNA gene sites in diploid and polyploid *Fragaria* (Rosaceae). *BMC Plant Biology* **11**: 157-169.

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.

Liu B, Wendel JF. 2002. Non-Mendelian phenomena in allopolyploid genome evolution. *Current genomics* **3**: 1-17.

Liu B, Wendel JF. 2003. Epigenetic phenomena and the evolution of plant allopolyploids. *Molecular Phylogenetics and Evolution* **29**: 365-379.

Lukens LN, Pires JC, Leon E, Vogelzang R, Oslach L, Osborn T. 2006. Patterns of sequence loss and cytosine methylation within a population of newly resynthesized *Brassica napus* allopolyploids. *Plant Physiology* **140**: 336-348.

Luo M, Yu C-W, Chen F-F, Zhao L, Tian G, Liu X, Cui Y, Yang J-Y, Wu K. 2012. Histone deacetylase HDA6 is functionally associated with AS1 in repression of *KNOX* genes in *Arabidopsis*. *PLoS Genetics* **8**: e1003114.

Lutz AM. 1907. A preliminary note on the chromosomes of *Oenothera lamarkiana* and one of its mutants *O. gigas. Science* **26**: 151-152.

Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* **290**: 1151-1155.

Ma X, Lv S, Zhang C, Yang C. 2013. Histone deacetylases and their functions in plants. *Plant Cell Reports* **32**: 465-478.

Ma XF, Fang P, Gustafson JP. 2004. Polyploidization-induced genome variation in Triticale. *Genome* **47**: 839-848.

Ma XF, Gustafson JP. 2008. Allopolyploidization-accommodated genomic sequence changes in Triticale. *Annals of Botany* **101**: 825-832.

Mable BK, Alexandrou MA, Taylor MI. 2011. Genome duplication in amphibians and fish: an extended synthesis. *Journal of Zoology* **284**: 151-182.

Madlung A. 2013. Polyploidy and its effect on evolutionary success: old questions revisited with new tools. *Heredity* **110**: 99-104.

Madlung A, Masuelli RW, Watson B, Reynolds SH, Davison J, Comai L. 2002. Remodeling of DNA methylation and phenotypic and transcriptional changes in synthetic *Arabidopsis* allotetraploids. *Plant Physiology* **129**: 733-746.

Madlung A, Tyagi AP, Watson B, Jiang HM, Kagochi T, Doerge RW, Martienssen R, Comai L. 2005. Genomic changes in synthetic *Arabidopsis* polyploids. *Plant Journal* **41**: 221-230.

Malinska H, Tate JA, Matyasek R, Leitch AR, Soltis DE, Soltis PS, Kovarik A. 2010. Similar patterns of rDNA evolution in synthetic and recently formed natural populations of *Tragopogon (*Asteraceae) allotetraploids. *BMC Evolutionary Biology* **10**: 291.

Malinska H, Tate JA, Mavrodiev E, Matyasek R, Lim KY, Leitch AR, Soltis DE, Soltis PS, Kovarik A. 2011. Ribosomal RNA genes evolution in *Tragopogon*: A story of new and old world allotetraploids and the synthetic lines. *Taxon* **60**: 348-354.

Mandel MA, Gustafson-Brown C, Savidge B, Yanofsky MF. 1992. Molecular characterization of the *Arabidopsis* floral homeotic gene *APETALA1*. *Nature* **360**: 273-277.

Mandel MA, Yanofsky MF. 1995. A gene triggering flower formation in *Arabidopsis*. *Nature* **377**: 522-524.

Manzara T, Gruissem W. 1988. Organization and expression of the genes encoding ribulose-1,5-bisphosphate carboxylase in higher-plants. *Photosynthesis Research* **16**: 117-139.

Martienssen RA, Colot V. 2001. DNA methylation and epigenetic inheritance in plants and filamentous fungi. *Science* **293**: 1070-1074.

Martin SL, Husband BC. 2012. Whole genome duplication affects evolvability of flowering time in an autotetraploid plant. *Plos One* **7**: e44784.

Mason A, Nelson M, Yan G, Cowling W. 2011. Production of viable male unreduced gametes in *Brassica* interspecific hybrids is genotype specific and stimulated by cold temperatures. *BMC Plant Biology* **11**: 103.

Matsushita SC, Tyagi AP, Thornton GM, Pires JC, Madlung A. 2012. Allopolyploidization lays the foundation for evolution of distinct populations: evidence from analysis of synthetic *Arabidopsis* allohexaploids. *Genetics* **191**: 535.

Matyasek R, Tate J, Lim Y, Srubarova H, Koh J, Leitch A, Soltis D, Soltis P, Kovarik A. 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, Mark Tancig, Anna M. Sherwood, Matthew A. Gitzendanner, Jennifer Rocca, Pamela S. Soltis, Douglas E. Soltis. 2005. Phylogeny of *Tragopogon L.* (Asteraceae) based on internal and external transcribed spacer sequence data. *International Journal of Plant Sciences* **166**: 117-133.

Melamed-Bessudo C, Levy AA. 2012. Deficiency in DNA methylation increases meiotic crossover rates in euchromatic but not in heterochromatic regions in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* **109**: E981-E988.

Mestiri I, Chague V, Tanguy AM, Huneau C, Huteau V, Belcram H, Coriton O, Chalhoub B, Jahier J. 2010. Newly synthesized wheat allohexaploids display progenitor-dependent meiotic stability and aneuploidy but structural genomic additivity. *New Phytologist* **186**: 86-101.

Meyerowitz EM, Bowman JL, Brockman LL, Drews GN, Jack T, Sieburth LE, Weigel D. 1991. A genetic and molecular model for flower development in Arabidopsis thaliana. *Development (Cambridge, England). Supplement* **1**: 157-167.

Miko I. 2008. Non-nuclear genes and their inheritance. *Nature Education* **1**.

Miziorko HM, Lorimer GH. 1983. Ribulose-1,5-bisphosphate carboxylase-oxygenase. *Annual Review of Biochemistry* **52**: 507-535.

Mlinarec J, Satovic Z, Malenica N, Ivancic-Bace I, Besendorfer V. 2012. Evolution of the tetraploid Anemone multifida (2n = 32) and hexaploid A. baldensis (2n = 48) (Ranunculaceae) was accompanied by rDNA loci loss and intergenomic translocation: evidence for their common genome origin. *Annals of Botany* **110**: 703-712.

Mocellin S, Provenzano M. 2004. RNA interference: learning gene knock-down from cell physiology. *Journal of Translational Medicine* **2**: 39.

Modliszewski JL, Willis JH. 2014. Near-absent levels of segregational variation suggest limited opportunities for the introduction of genetic variation via homeologous chromosome pairing in synthetic neoallotetraploid *mimulus*. *G3 (Bethesda, Md.)* **4**: 509-522.

Mun J, Kwon S, Yang T, Seol Y, Jin M, Kim Jin A, Lim M, Kim J, Baek S, Choi B, *et al.* 2009. Genome-wide comparative analysis of the *Brassica rapa* gene space reveals genome shrinkage and differential loss of duplicated genes after whole genome triplication. *Genome Biology* **10**: R111-R111.

Munster T, Wingen LU, Faigl W, Werth S, Saedler H, Theissen G. 2001. Characterization of three *GLOBOSA*-like MADS-box genes from maize: evidence for ancient paralogy in one class of floral homeotic B-function genes of grasses. *Gene* **262**: 1-13.

Myers WM, Hill HD. 1940. Studies of chromosomal association and behavior and occurrence of aneuploidy in autotetraploid grass species, orchard grass, tall oat grass, and crested wheat grass. *Botanical Gazette* **102**: 236-255.

Neff MM, Turk E, Kalishman M. 2002. Web-based primer design for single nucleotide polymorphism analysis. *Trends in Genetics* **18**: 613-615.

Ng DWK, Lu J, Chen ZJ. 2012. Big roles for small RNAs in polyploidy, hybrid vigor, and hybrid incompatibility. *Current Opinion in Plant Biology* **15**: 154-161.

Ng M, Yanofsky MF. 2000. Three ways to learn the ABCs. *Curr Opin Plant Biol* **3**: 47-52.

Ng M, Yanofsky MF 2001. Function and evolution of the plant MADS-BOX gene family.In. *Nature Reviews Genetics*: Nature Publishing Group. 186-195.

Nicolas SD, Le Mignon G, Eber F, Coriton O, Monod H, Clouet V, Huteau V, Lostanlen A, Delourme R, Chalhoub B, *et al.* 2007. Homeologous recombination plays a major role in chromosome rearrangements that occur during meiosis of *Brassica napus* haploids. *Genetics* **175**: 487-503.

Nicolas SD, Monod H, Eber F, Chevre A-M, Jenczewski E 2012. Non-random distribution of extensive chromosome rearrangements in *Brassica napus* depends on genome organization.In. *Plant Journal*: Wiley-Blackwell. 691-703.

Oehlkers F. 1964. Cytoplasmic inheritance in the genus *Streptocarpus lindley*. *Advances in Genetics* **12**: 329-370.

Olmstead RG, Micheals HJ, Scott KM, Palmer JD. 1992. Monophyly of the Asteridae and identification of their major lineages inferred from DNA sequences of *rbcL*. *Annals of the Missouri Botanical Garden* **79**: 249-265.

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

Otto S. 2007. The evolutionary consequences of polyploidy. *Cell* **131**: 452 - 462.

Outchkourov NS, Peters J, Jong J, Rademakers W, Jongsma MA. 2003. The promoter-terminator of chrysanthemum *rbcS1* directs very high expression levels in plants. *Planta* **216**: 1003-1012.

Ownbey M. 1950. Natural hybridization and amphiploidy in the genus *Tragopogon*. *American Journal of Botany* **37**: 487-499.

Ownbey M, McCollum GD. 1953. Cytoplasmic inheritance and reciprocal amphiploidy in *Tragopogon*. *American Journal of Botany* **40**: 788-796.

Ownbey M, McCollum GD. 1954. The chromosomes of *Tragopogon*. *Rhodora* **56**: 7-21.

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

Parida A, Raina SN, Narayan RKJ. 1990. Quantitative DNA Variation between and within Chromosome Complements of *Vigna* Species (Fabaceae). *Genetica* **82**: 125-133.

Parisod C, Alix K, Just J, Petit M, Sarilar V, Mhiri C, Ainouche M, Chalhoub B, Grandbastien MA. 2010. Impact of transposable elements on the organization and function of allopolyploid genomes. *New Phytologist* **186**: 37-45.

Parisod C, Salmon A, Zerjal T, Tenaillon M, Grandbastien MA, Ainouche M. 2009. Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in *Spartina*. *New Phytol* **184**: 1003-1015.

Paterson AH, Bowers JE, Chapman BA. 2004. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 9903-9908.

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 in Genetics* **22**: 597-602.

Patterson B. 2009. Systematics, evolution, and biogeography of compositae. *Madron* **56**: 209-211.

Paun O, Bateman R, Fay M, Hedren M, Civeyrel L, Chase M. 2010. Stable epigenetic effects impact adaptation in allopolyploid orchids (Dactylorhiza: Orchidaceae). *Molec Biol Evol* **27**: 2465 - 2473.

Paun O, Bateman R, Fay M, Luna J, Moat J, Hedren M, Chase M. 2011. Altered gene expression and ecological divergence in sibling allopolyploids of Dactylorhiza (Orchidaceae). *BMC Evolutionary Biology* **11**: 113.

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

Peakall R, Smouse PE. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**: 288-295.

Pecrix Y, Rallo G, Folzer H, Cigna M, Gudin S, Le Bris M. 2011. Polyploidization mechanisms: temperature environment can induce diploid gamete formation in *Rosa* sp. *Journal of Experimental Botany* **62**: 3587-3597.

Pelaz S, S. DG, Elvira B, Ellen W, F. YM. 2000. B and C floral organ identity functions require *SEPALLATA* MADS-box genes. *Nature* **405**: 200.

Pellicer J, Garnatje T, Hidalgo O, Tagashira N, Valles J, Kondo K. 2010. Do polyploids require proportionally less rDNA loci than their corresponding diploids? Examples from Artemisia subgenera Absinthium and Artemisia (Asteraceae, Anthemideae). *Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology* **144**: 841-848.

Pinto AV, Mathieu Al, Marsin Sp, Veaute X, Ielpi L, Labigne As, Radicella JP. 2005. Suppression of Homologous and Homeologous Recombination by the Bacterial MutS2 Protein. *Molecular Cell* **17**: 113-120.

Pinyopich A, Ditta GS, Savidge B, Liljegren SJ, Baumann E, Wisman E, Yanofsky MF. 2003. Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature* **424**: 85-88.

Pires JC, Lim KY, Kovarik A, Matyasek R, Boyd A, Leitch AR, Leitch IJ, Bennett MD, Soltis PS, Soltis DE. 2004a. Molecular cytogenetic analysis of recently evolved *Tragopogon* (Asteraceae) allopolyploids reveal a karyotype that is additive of the diploid progenitors. *American Journal of Botany* **91**: 1022-1035.

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

Pont C, Murat F, Confolent C, Balzergue S, Salse J. 2011. RNA-seq in grain unveils fate of neo- and paleopolyploidization events in bread wheat (*Triticum aestivum L.*). *Genome Biology* **12**.

Pontes O, Neves N, Silva M, Lewis MS, Madlung A, Comai L, Viegas W, Pikaard CS. 2004. Chromosomal locus rearrangements are a rapid response to formation of the allotetraploid Arabidopsis suecica genome. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 18240-18245.

Prince VE, Pickett FB. 2002. Splitting pairs: the diverging fates of duplicated genes. *Nature Reviews Genetics* **3**: 827-837.

Ramanna MS. 1979. A re-examination of the mechanisms of 2n gamete formation in potato and its implications for breeding. *Euphytica* **28**: 537-561.

Ramsey J, Schemske D. 2002. Neopolyploidy in flowering plants. *Annu. Rev. Ecol. Syst.* **33**: 589–639.

Ramsey J, Schemske DW. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics* **29**: 467-501.

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

Renny-Byfield S, Chester M, Kovarik A, Le Comber SC, Grandbastien M-A, Deloger M, Nichols RA, Macas J, Novak P, Chase MW, *et al.* 2011. Next generation sequencing reveals genome downsizing in allotetraploid *Nicotiana tabacum*, predominantly through the elimination of paternally Derived repetitive DNAs. *Molecular Biology and Evolution* **28**: 2843-2854.

Renny-Byfield S, Kovarik A, Kelly LJ, Macas J, Novak P, Chase MW, Nichols RA, Pancholi MR, Grandbastien M-A, Leitch AR. 2013. Diploidization and

genome size change in allopolyploids is associated with differential dynamics of low- and high-copy sequences. *The Plant Journal* **74**: 829-839.

Richards CL, Walls RL, Bailey JP, Parameswaran R, George T, Pigliucci M. 2008. Plasticity in salt tolerance traits allows for invasion of novel habitat by Japanese knotweed s. l. (Fallopia japonica and F-bohemica, Polygonaceae). *American Journal of Botany* **95**: 931-942.

Riechmann JL, Meyerowitz EM. 1997. MADS domain proteins in plant development. *Biological Chemistry* **378**: 1079-1101.

Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nature Biotechnology* **29**: 24-26.

Rodermel SR, Abbott MS, Bogorad L. 1988. Nuclear-organelle interactions - nuclear antisense gene inhibits ribulose bisphosphate carboxylase enzyme levels in transformed tobacco plants. *Cell* **55**: 673-681.

Roose ML, Gottlieb LD. 1976. Genetic and biochemical consequences of polyploidy in *Tragopogon*. *Evolution* **30**: 818-830.

Roulin A, Auer PL, Libault M, Schlueter J, Farmer A, May G, Stacey G, Doerge RW, Jackson SA. 2013. The fate of duplicated genes in a polyploid plant genome. *The Plant Journal* **73**: 143-153.

Rozen S, Skaletsky HJ 2002. Primer3 on the WWW for general users and for biologist programmers. In: S. KrawetzS. Misener eds. *Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, USA*, 365-386.

Russell A, Safer S, Weiss-Schneeweiss H, Temsch E, Stuppner H, Stuessy TF, Samuel R. 2013. Chromosome counts and genome size of *Leontopodium* species (Asteraceae: Gnaphalieae) from south-western China. *Botanical Journal of the Linnean Society* **171**: 627-636.

Saedler H, Huijser P. 1993. Molecular biology of flower development in *Antirrhinum majus* (snapdragon). *Gene* **135**: 239-243.

Salmon A, Ainouche ML 2010. Polyploidy and DNA methylation: new tools available.In. *Molecular Ecology*: Wiley-Blackwell. 213-215.

Salmon A, Ainouche ML, Wendel JF. 2005. Genetic and epigenetic consequences of recent hybridization and polyploidy in *Spartina* (Poaceae). *Molecular Ecology* **14**: 1163-1175.

Salmon A, Clotault Jrm, Jenczewski E, Chable Vr, Manzanares-Dauleux MJ. 2008. *Brassica oleracea* displays a high level of DNA methylation polymorphism. *Plant Science* **174**: 61-70.

Sankoff D, Nadeau J, Vision T, Brown D 2000. Genome archaeology: detecting ancient polyploidy in contemporary genomes. *Comparative Genomics*. Netherlands: Springer 479-491.

Sasanuma T. 2001. Characterization of the *rbcS* multigene family in wheat: subfamily classification, determination of chromosomal location and evolutionary analysis. *Molecular Genetics and Genomics* **265**: 161-171.

Sax K. 1936. The experimental production of polyploidy. *J. Arnold Arbor* **17**: 153-159.

Schuelke M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnololgy* **18**: 233-234.

Sears ER. 1976. Genetic-control of chromosome-pairing in wheat. *Annual Review of Genetics* **10**: 31-51.

Seoighe C, Gehring C. 2004. Genome duplication led to highly selective expansion of the Arabidopsis thaliana proteome. *Trends Genet* **20**: 461 - 464.

Severin AJ, Cannon SB, Graham MM, Grant D, Shoemaker RC. 2011. Changes in twelve homoeologous genomic regions in soybean following three rounds of polyploidy. *The Plant Cell* **23**: 3129-3136.

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.

Shawn JC, Suhua F, Xiaoyu Z, Zugen C, Barry M, Christian DH, Sriharsa P, Stanley FN, Matteo P, Steven EJ. 2008. Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* **452**: 215-219.

Shchennikova AV, Shulga OA, Immink R, Skryabin KG, Angenent GC. 2004. Identification and characterization of four *Chrysanthemum* MADS-box genes, belonging to the *APETALA1*/*FRUITFULL* and *SEPALLATA3* Subfamilies. *Plant Physiology* **134**: 1632-1641.

Shirsat A, Wilford N, Croy R, Boulter D. 1989. Sequences responsible for the tissue specific promoter activity of a pea legumin gene in tobacco. *Molecular & General Genetics* **215**: 326-331.

Shul'ga OA, Shennikova AV, Angenent GS, Skriabin KG. 2008. MADS-box genes controlling inflorescence morphogenesis in sunflower. *Russian Journal of Developmental Biology* **39**: 2-5.

Siebert PD, Chenchik A, Kellogg DE, Lukyanov KA, Lukyanov SA. 1995. An improved PCR method for walking in uncloned genomic DNA. *Nucleic Acids Research* **23**: 1087-1088.

Skalicka K, Lim KY, Matyasek R, Matzke M, Leitch AR, Kovarik A. 2005. Preferential elimination of repeated DNA sequences from the paternal, *Nicotiana tomentosiformis* genome donor of a synthetic, allotetraploid tobacco. *New Phytologist* **166**: 291-303.

Smith ZD, Gu H, Bock C, Gnirke A, Meissner A. 2009. High-throughput bisulfite sequencing in mammalian genomes. *Methods* **48**: 226-232.

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

Soltis DE, Chanderbali AS, Kim S, Buzgo M, Soltis PS. 2007. The ABC model and its applicability to basal angiosperms. *Annals of Botany* **100**: 155-163.

Soltis DE, Leebens-Mack JH, Soltis PS. 2006. *Developmental genetics of the flower.* Amsterdam, Boston Academic Press.

Soltis DE, Soltis PS. 1989. Allopolyploid speciation in *Tragopogon* - insights from chloroplast DNA. *American Journal of Botany* **76**: 1119-1124.

Soltis DE, Soltis PS. 1995. The dynamic nature of polyploid genomes. *Proceedings of the National Academy of Sciences of the United States of America* **92**: 8089-8091.

Soltis DE, Soltis PS. 1999. Polyploidy: recurrent formation and genome evolution. *Trends in Ecology and Evolution* **14**: 348-352.

Soltis DE, Soltis PS, Pires JC, Kovarik A, Tate JA, Mavrodiev E. 2004. Recent and recurrent polyploidy in *Tragopogon* (Asteraceae): cytogenetic, genomic and genetic comparisons. *Biological Journal of the Linnean Society* **82**: 485-501.

Soltis PS, Plunkett GM, Novak SJ, Soltis DE. 1995a. Genetic-variation in *Tragopogon* species - additional origins of the allotetraploids *T. mirus* and *T. miscellus* (Compositae). *American Journal of Botany* **82**: 1329-1341.

Soltis PS, Plunkett GM, Novak SJ, Soltis DE. 1995b. Genetic variation in *Tragopogon* species: additional origins of the allotetraploids *T. mirus* and *T. miscellus* (Compositae). *American Journal of Botany* **82**: 1329-1341.

Soltis PS, Soltis DE. 1991. Multiple origins of the allotetraploid *Tragopogon mirus* (Compositae) - rDNA evidence. *Systematic Botany* **16**: 407-413.

Soltis PS, Soltis DE. 2000. The role of genetic and genomic attributes in the success of polyploids. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 7051-7057.

Soltis PS, Soltis DE. 2004. The origin and deversification of angiosperms. *American Journal of Botany* **91**: 1614-1626.

Soltis PS, Soltis DE 2009. The role of hybridization in plant speciation. *Annual Review of Plant Biology*, 561-588.

Sommer H, Beltran JP, Huijser P, Pape H, Lonnig WE, Saedler H, Schwarzsommer Z. 1990. *DEFICIENS*, a homeotic gene involved in the control of flower morphogenesis in *Antirrhinum majus* - the protein shows homology to transcription factors. *Embo Journal* **9**: 605-613.

Song C-F, Lin Q-B, Liang R-H, Wang Y-Z. 2009. Expressions of ECE-*CYC2* clade genes relating to abortion of both dorsal and ventral stamens in *Opithandra* (Gesneriaceae). *BMC Evolutionary Biology* **9**: 244.

Song KM, Lu P, Tang KL, Osborn TC. 1995. Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. *Proceedings of the National Academy of Sciences of the United States of America* **92**: 7719-7723.

Spreitzer RJ. 2003. Role of the small subunit in ribulose-1,5-bisphosphate carboxylase/oxygenase. *Archives of Biochemistry and Biophysics* **414**: 141-149.

Spreitzer RJ, Salvucci ME. 2002. Rubisco: Structure, regulatory interactions, and possibilities for a better enzyme. *Annual Review of Plant Biology* **53**: 449-475.

Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**: 637-644.

Staswick PE. 1994. Storage proteins of vegetative plant tissues. *Annual Review of Plant Physiology and Plant Molecular Biology* **45**: 303-322.

Stebbins GL 1947. Types of Polyploids: Their Classification and Significance. *Advances in Genetics*: Academic Press, 403-429.

Stebbins GL. 1950. Variation and evolution in plants. *Columbia University Press, New York*.

Stubbe H 1966. Genetics and Cytology of the Genus *Antirrhinum* section *Antirrhinum*. VEB Gustav Fischer Verlag: Jena, 421.

Swaminathan MS. 1954. Nature of polyploidy in some 48-chromosome species of the genus *Solanum*, section *tuberarium*. *Genetics* **39**: 59-76.

Sybenga J. 1996. Chromosome pairing affinity and quadrivalent formation in polyploids: do segmental allopolyploids exist? *Genome* **39**: 1176-1184.

Symonds VV, Lloyd AM. 2004. A simple and inexpensive method for producing fluorescently labelled size standard. *Molecular Ecology Notes* **4**: 768–771.

Symonds VV, Soltis PS, Soltis DE. 2010. Dynamics of polyploid formation in *Tragopogon* (Asteraceae): recurrent formation, gene flow, and population structure. *Evolution* **64**: 1984-2003.

Tang ZX, Wu M, Zhang HQ, Yan BJ, Tan FQ, Zhang HY, Fu SL, Ren ZL. 2012. Loss of parental coding sequences in an early generation of wheat-rye allopolyploid. *International Journal of Plant Sciences* **173**: 1-6.

Tate J, Ni Z, Scheen A, Koh J, Gilbert C, Lefkowitz D, Chen Z, Soltis P, Soltis D. 2006a. Evolution and expression of homeologous loci in Tragopogon miscellus (Asteraceae), a recent and reciprocally formed allopolyploid. *Genetics* **173**: 1599 - 1611.

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

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

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

Tate JA, Soltis DE, Soltis PS. 2005. Polyploidy in plants. In: The evolution of the plant genome. *Elsevier Inc.* : 371-426.

Tate JA, Symonds VV, Doust AN, Buggs RJA, Mavrodiev E, Majure LC, Soltis PS, Soltis DE. 2009b. Synthetic polyploids of *Tragopogon miscellus* and *T. mirus* (Asteraceae): 60 years after Ownbey's discovery. *American Journal of Botany* **96**: 979-988.

Te Beest M, Le Roux JJ, Richardson DM, Brysting AK, Suda J, Kubesova M, Pysek P. 2012. The more the better? The role of polyploidy in facilitating plant invasions. *Annals of Botany* **109**: 19-45.

Thompson JD, Lumaret R. 1992. The evolutionary dynamics of polyploid plants: origins, establishment and persistence. *Trends in Ecology & Evolution* **7**: 302-307.

Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics* **14**: 178-192.

Trobner W, Ramirez L, Motte P, Hue I, Huijser P, Lonnig WE, Saedler H, Sommer H, Schwarzsommer Z. 1992. *GLOBOSA* - a homeotic gene which interacts with *DEFICIENS* in the control of *Antirrhinum* floral organogenesis. *Embo Journal* **11**: 4693-4704.

Tsuchimoto S, Vanderkrol AR, Chua NH. 1993. Ectopic expression of *pMADS3* in transgenic *Petunia* phenocopies the *Petunia* blind mutant. *Plant Cell* **5**: 843-853.

Udall JA, Wendel JF. 2006. Polyploidy and crop improvement. *The plant Genome* **1**: S3-S14.

Uimari A, Kotilainen M, Elomaa P, Yu DY, Albert VA, Teeri TH. 2004. Integration of reproductive meristem fates by a *SEPALLATA*-like MADS-box gene. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 15817-15822.

Van Laere K, Franca SC, Vansteenkiste H, Van Huylenbroeck J, Steppe K, Van Labeke M-C. 2011. Influence of ploidy level on morphology, growth and drought susceptibility in *Spathiphyllum wallisii*. *Acta Physiologiae Plantarum* **33**: 1149-1156.

Vandenbussche M, Zethof J, Souer E, Koes R, Tornielli GB, Pezzotti M, Ferrario S, Angenent GC, Gerats T. 2003. Toward the analysis of the *Petunia* MADS box gene family by reverse and forward transposon insertion mutagenesis

approaches: B, C, and d floral organ identity functions require *SEPALLATA*-like mads box genes in *Petunia*. *Plant Cell* **15**: 2680-2693.

Vanderkrol AR, Chua NH. 1993. Flower development in *Petunia*. *Plant Cell* **5**: 1195-1203.

Vanyushin BF 2006. DNA methylation in plants. In: W. DoerflerP. Böhm eds. *DNA Methylation: Basic Mechanisms*. Berlin Heidelberg: Springer 67-122.

Vanyushin BF, Ashapkin VV. 2011. DNA methylation in higher plants: past, present and future. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1809**: 360-368.

Vaughan TW. 1906. The work of Hugo de Vries and its importance in the study of problems of evolution. *Science* **23**: 681-691.

Verhoeven KJ, Van Dijk PJ, Biere A. 2010. Changes in genomic methylation patterns during the formation of triploid asexual dandelion lineages. *Molecular Ecology* **19**: 315-324.

Vincze T, Posfai J, Roberts RJ. 2003. NEBcutter: a program to cleave DNA with restriction enzymes. *Nucleic Acids Research* **31**: 3688-3691.

Waksman G, Lebrun M, Freyssinet G. 1987. Nucleotide-sequence of a gene encoding sunflower ribulose-1,5-bisphosphate carboxylase oxygenase small subunit (*rbcS*). *Nucleic Acids Research* **15**: 7181-7181.

Wang A, Zhang X, Yang C, Song Z, Du C, Chen D, He Y, Cai D. 2013a. Development and characterization of synthetic amphiploid (AABB) between *Oryza sativa* and *Oryza punctata*. *Euphytica* **189**: 1-8.

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

Wang X, Shi X, Hao B, Ge S, Luo J. 2005. Duplication and DNA segmental loss in the rice genome: implications for diploidization. *New Phytologist* **165**: 937-946.

Wang Y, Wang X, Lee T-H, Mansoor S, Paterson AH. 2013b. Gene body methylation shows distinct patterns associated with different gene origins and duplication modes and has a heterogeneous relationship with gene expression in *Oryza sativa* (rice). *New Phytologist* **198**: 274-283.

Wang Y, Wang X, Tang H, Tan X, Ficklin SP, Feltus FA, Paterson AH. 2011. Modes of gene duplication contribute differently to genetic novelty and redundancy, but show parallels across divergent angiosperms. *Plos One* **6**: e28150.

Warnecke PM, Stirzaker C, Song J, Grunau C, Melki JR, Clark SJ. 2002. Identification and resolution of artifacts in bisulfite sequencing. *Methods* **27**: 101-107.

Weber M, Schuebeler D. 2007. Genomic patterns of DNA methylation: targets and function of an epigenetic mark. *Current Opinion in Cell Biology* **19**: 273-280.

Weigel D, Alvarez J, Smyth DR, Yanofsky MF, Meyerowitz EM. 1992. *LEAFY* controls floral meristem identity in *Arabidopsis*. *Cell* **69**: 843-859.

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

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

Wendel JF, Flagel LE, Adams KL 2012. Jeans, genes, and genomes: cotton as a model for studying polyploidy. In: P. S. SoltisD. E. Soltis eds. *Polyploidy and genome evolution*. Berlin Heidelberg: Springer 181-207.

Werner JE, Peloquin SJ. 1991. Occurrence and mechanisms of 2n egg formation in 2x potato. *Genome* **34**: 975-982.

Whipple CJ, Schmidt RJ, Douglas E. Soltis JHLPSS, Callow JA 2006. Genetics of grass flower development. *Advances in botanical research*: Academic Press, 385-424.

Wolf JB. 2009. Cytonuclear interactions can favor the evolution of genomic imprinting. *Evolution* **63**: 1364-1371.

Wolfe KH. 2001. Yesterday's polyploids and the mystery of diploidization. *Nature Reviews Genetics* **2**: 333-341.

Wong C, Murray BG. 2012. Variable Changes in Genome Size Associated with Different Polyploid Events in *Plantago* (Plantaginaceae). *Journal of Heredity* **103**: 711-719.

Wright KM, Pires JC, Madlung A. 2009. Mitotic instability in resynthesized and natural polyploids of the genus *Arabidopsis* (Brassicaceae). *American Journal of Botany* **96**: 1656-1664.

Wu RL, Gallo-Meagher M, Littell RC, Zeng ZB. 2001. A general polyploid model for analyzing gene segregation in outcrossing tetraploid species. *Genetics* **159**: 869-882.

Xiong Z, Gaeta RT, Pires JC. 2011. Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. *Proceedings of the National Academy of Sciences* **108**: 7908-7913.

Xu Y, Xu H, Wu X, Fang X, Wang J. 2012. Genetic changes following hybridization and genome doubling in synthetic *Brassica napus*. *Biochemical Genetics* **50**: 616-624.

Yaakov B, Kashkush K. 2011. Methylation, transcription, and rearrangements of transposable elements in synthetic allopolyploids. *International Journal of Plant Genomics* **2011**: 7.

Yang XH, Ye CY, Cheng ZM, Tschaplinski TJ, Wullschleger SD, Yin WL, Xia XL, Tuskan GA. 2011a. Genomic aspects of research involving polyploid plants. *Plant Cell Tissue and Organ Culture* **104**: 387-397.

Yang Y, Zhang L, Yan Z, Zheng Y, Liu D. 2011b. The cytological instability of neoallopolyploids suggesting a potent way for DNA introgression: The case of synthetic hexaploid wheat x *Aegilops peregrina*. *African Journal of Agricultural Research* **6**: 1692-1697.

Yanofsky MF, Ma H, Bowman JL, Drews GN, Feldmann KA, Meyerowitz EM. 1990. The protein encoded by the *Arabidopsis* homeotic gene *AGAMOUS* resembles transcription factors. *Nature* **346**: 35-39.

Yoo MJ, Szadkowski E, Wendel JF. 2013. Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity* **110**: 171-180.

Yu DY, Kotilainen M, Pollanen E, Mehto M, Elomaa P, Helariutta Y, Albert VA, Teeri TH. 1999. Organ identity genes and modified patterns of flower development in *Gerbera hybrida* (Asteraceae). *Plant Journal* **17**: 51-62.

Yu Z, Haberer G, Matthes M, Rattei T, Mayer KFX, Gierl A, Torres-Ruiz RA. 2010. Impact of natural genetic variation on the transcriptome of autotetraploid *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 17809-17814.

Zeng ZX, Zhang T, Li GR, Liu C, Yang ZJ. 2012. Phenotypic and epigenetic changes occurred during the autopolyploidization of *Aegilops tauschii*. *Cereal Research Communications* **40**: 476-485.

Zhang X, Ge X, Shao Y, Sun G, Li Z. 2013. Genomic change, retrotransposon mobilization and extensive cytosine methylation alteration in *Brassica napus* introgressions from two intertribal hybridizations. *Plos One* **8**.

Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SWL, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, *et al.* 2006. Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell* **126**: 1189-1201.

Zielinski M-L, Scheid OM 2012. Meiosis in polyploids. In: D. E. SoltisP. S. Soltis eds. *Polyploidy and genome evolution*. Berlin Heidelberg: Springer, 33-55.

Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S. 2007. Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nature Genetics* **39**: 61-69.