Evolution of Cytonuclear Coordination in *Tragopogon* (Asteraceae) Allopolyploids

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

Doctor of Philosophy

in

Plant Biology

at Massey University, Manawatu, New Zealand

Sidra Hussain

2023

Abstract

Cytonuclear coordination is an important aspect in plant evolution which involves the synchronized interactions between nuclear and organellar genomes. Allopolyploidy, resulting from interspecific hybridization and genome duplication, could result in cytonuclear incompatibilities. Therefore, to coordinate cytonuclear interactions, allopolyploids may undergo alterations in duplicated nuclear gene expression via incorporating maternally biased expression patterns. To investigate cytonuclear responses to allopolyploidy, in this study, expression patterns of duplicated nuclear genes and their organelle counterparts, implicated in cytonuclear enzyme complexes, as well as dual-targeted genes were investigated in the reciprocally formed young *T. miscellus* allopolyploids (90-100 years) and diploid parent species; *T. dubius* and *T. pratensis*. In addition, the effect of polyploidy on morphological traits of *T. miscellus* allopolyploids were examined and assessed relative to parent species. The expression data showed that *T. miscellus* allopolyploids are regulating expression at the homeolog level, primarily through *T. pratensis* bias, while maintaining the total gene expression levels as to parental levels. The morphological evaluation of allopolyploids and diploids demonstrated that both reciprocal forms of *T. miscellus* have significantly longer, but fewer leaves compared to the diploid parent species. These findings reflect that young *T. miscellus* allopolyploids are certainly undergoing homeolog expression regulation to accommodate cytonuclear interactions as well as displaying morphological responses to allopolyploidy. This study provides insights into polyploid genome evolution and contributes to further understanding of the cytonuclear coordination in allopolyploids.

Acknowledgements

First and foremost, all the praises belong to Allah Azawajal, who is Mighty and Majestic. Without Him being in my heart, I would not have embarked on this long journey.

I then, primarily, would like to enunciate my sincere gratitude and indebtedness to my Supervisors, Associate Professor Jennifer A. Tate, Professor Murray Cox, and Dr. David Winter for their constant support in my Ph.D. studies. It is a privilege to have such advisers and mentors whose extensive knowledge in the domain not only helped me in my studies but also in my professional development. I could not thank them enough for their patience, motivation, and guidance throughout the period of my doctoral journey.

Besides my advisers, the team of Massey Genome Service (MGS), NeSI, and Massey Plant Growth Unit has provided me with a great support. Special thanks to Xiaoxiao Lin and Anja Schiemann for their insightful comments and encouragement during the journey of the thesis. It not only provoked me to widen my research but also their support in learning was admirable.

Also, my sincere thanks goes to the LoST lab group which provided me with the opportunity to join the team as a researcher. Especially grateful to Thilini, Sofie, Kay, Weixuan and Usama for their inestimable support and patience. They always stimulated discussions and ideas to mull over.

Apart from the research team, I would like to thank my close friends exclusively Mustafa Kamal Pasha, Umm-e-Aiman, Dr Sehrish Nawaz, Saira Rao and Nadia Latif, whose kind support has always ensured me reach my true potential. My special thanks go to some unforgettable people I've met in this journey, Thilini, Maria, Amat-ur-Rahman, Atika Khurram, Nida Haroon and Tina Sehrish, who have given me reasons to continue on this journey. Without their support, it would have been a difficult task to complete.

Last but not the least, my gratitude goes to my family: my parents, Zubaida Akhtar and Ali Hussain; brother, Gulraiz Hussain; sisters, Aqsa Hussain and Ayesha Latif; late auntie, Shamim Akhtar; uncles, Syed Zulfiqar, Javaid, Faiz; nephews, Ibrahim and Ismail; niece, Hafsa; and late grandparents for continuous encouragement, prayers, and unshakeable belief. Their spiritual and moral support throughout this phase will truly be remembered in my life.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AltA	Alternative acceptor
AltD	Alternative donor
AltFE	Alternative first exon
AltLE	Alternative last exon
ANOVA	Analysis of variance
AS	Alternative splicing
AtIPT3	Adenosine phosphate-isopentenyl-transferase
BLAST	Basic local alignment search tool
Clp	Caseinolytic protease
CyMIRA	The Cytonuclear Molecular Interactions Reference for Arabidopsis
DCL	Dicer like proteins
DNA	Deoxyribonucleic acid
ERCC	External RNA Controls Consortium
ES	Exon skipping
FLC	Flowering locus C
GFP	Green fluorescent protein
GR	Genetic recombination
IR	Intron retention
LSU	Large subunit
MAFFT	Multiple alignment using fast fourier transform
MI	Meiosis I
MII	Meiosis II
mRNA	Messenger ribonucleic acid
My	Million years
NGE	Nuclear gene expression
NMD	Nonsense-mediated decay
Nt	Nucleotide

OGE	Organelle gene expression
OXPHOS	Oxidative phosphorylation
PAS2	PASTICCINO 2
PCA	Principal component analysis
PCR	Polymerace chain reaction
PTCs	Premature termination codons
PTMs	Post-translational modifications
QC	Quality control
RdDM	RNA directed DNA methylation
RNA	Ribonucleic acid
RNAi	Ribonucleic acid interference
RNA-Seq	Ribonucleic acid sequencing
ROS	Reactive oxygen species
RQS	Ribonucleic acid quality score
rRNA	Ribosomal ribonucleic acid
RT-PCR	Reverse transcription polymerase chain reaction
siRNA	Small interfering ribonucleic acid
smRNA	Small ribonucleic acid
SNP	Single nucleotide polymorphism
SSU	Small subunit
STAR	Spliced transcripts alignment to a reference
Tat	Twin-arginine translocation
TE	Transposable element
TF	Transcription factor
tRNA	Transfer ribonucleic acid
ul	Microliter
uM	Micromolar
USA	Unites States of America
WGD	Whole genome duplications

Chapter 1

Introduction

1.1. Polyploidy

Polyploidy is characterized by the presence of more than two complete sets of chromosomes within an organism. Polyploidization is considered to be a significant mechanism of plant speciation and diversification, thus contributing a major part in plant evolution (Otto and Whitton 2000; Adams and Wendel 2005; Soltis et al. 2009; Madlung 2013a). Although polyploidy has occurred in animals such as insects, reptiles, fishes and amphibians (reviewed in Van de Peer et al. 2017a), it is more common in plants (Otto and Whitton 2000). Many plant lineages have experienced one or more rounds of ancient polyploidization events through their evolutionary history (De Bodt et al. 2005; Cui et al. 2006; Soltis et al. 2009). For instance, genomic data revealed that all angiosperms, at least once in their evolutionary history, encountered ancient whole genome duplications (Jiao et al. 2011b). Additionally, recent polyploidy events have been found in almost 25-35% plant species (Salman-Minkov et al. 2016b). In support of this notion, many studies have identified the presence of repeated rounds of whole genome duplications during the course of plant evolution and even species with small genomes that were long believed to be diploids, like *Arabidopsis thaliana,* have experienced ancient polyploidy (Adams and Wendel 2005). The frequency of polyploidy for different groups of green plants is wide-ranging, from very low in bryophytes to 31% in ferns and up to 15% in angiosperms (Wood et al. 2009). Polyploidization events are documented as key events in the history of seed plants (Fawcett et al. 2009; Jiao et al. 2011b; Leitch and Leitch 2013). Even within the angiosperm group, polyploidy is more frequent in monocots than dicots (Otto and Whitton 2000). Moreover, closely related species and even different populations within

species may have ploidy level differences and multiple origins as well (Lewis 1980; Soltis et al. 1993; Soltis and Soltis 1999). These observations propose that polyploidy in plants is a constant process leading to evolutionary innovation. As many important crops are recent polyploids, including wheat, sugarcane, potato, canola and cotton, or the outcome of an ancient polyploidization event, such as rice, maize, and soybean (Leitch and Leitch 2008).

1.2. Classification of Polyploids

Stebbins characterized three main types of polyploids; autopolyploids, segmental allopolyploids and allopolyploids (Stebbins Jr 1947) based on their origin and chromosome pairing behavior (Chen and Ni 2006b; Grant 1981b; Stebbins 1971b). However, in many cases, a species cannot be clearly designated as one type or another, leading to uncertainty or debate on the subject (Barker et al. 2016; Doyle and Sherman-Broyles 2017). Autopolyploids undergo the process of genome duplication within the same species, which results from either identical or quite similar genomes (Stebbins Jr 1947; Lewis 1980). In allopolyploids, hybridization between two different species followed by genome doubling, leads to two or more divergent genomes within the same nucleus (Stebbins Jr 1947; Grant 1978). Segmental allopolyploids involve the doubling of partially distinct genomes and include natural hybrid polyploids forming multivalents (Stebbins Jr 1947; Levin 2002). Furthermore, aneuploids, which either lack a chromosome or possess an extra one, may arise from meiotic errors leading to gametes with various haploid numbers (Myers and Hill 1940).

Chromosome pairing leads to distinct types of polyploids (Jackson 1982). Multivalent formation and polysomic inheritance play a crucial role in characterizing autopolyploids, whereas bivalent formation and disomic inheritance are the distinguished

features of allopolyploids (Stebbins 1971a; Ramsey and Schemske 2002; Comai 2005b; Le Comber et al. 2010; Parisod et al. 2010b; Zielinski and Scheid 2012). Exceptional patterns of inheritance are observed either in autopolyploids showing consistent bivalent formation (Weiss and Maluszynska 2001) or segmental allopolyploids that show a mixture of disomic and polysomic segregation patterns (Stebbins 1971a).

Based on the age of whole genome duplication (WGD) events, polyploids can be classified as neo-, meso- or paleopolyploids (Mandáková et al. 2010), but ideas around these ages can vary. Neopolyploids (~ < 200,000 years old) are newly synthesized young polyploids, which are characterized by clearly distinguishable progenitor sub-genomes and extant diploid ancestors (Ramsey and Schemske 2002; Kagale et al. 2014; Vallejo-Marín et al. 2015). With the passage of time, neopolyploids evolve into mesopolyploids (~ 8-17 My ago) whose sub-genomes are only discernible via comparative genetic and genomic approaches (Parkin et al. 2005; Mandáková et al. 2010; Parkin et al. 2014; Wang et al. 2011). The term paleopolyploids refers to polyploids that have diploidized over several million years through extensive genome restructuring that leads to the assimilation of parental sub-genomes (Wolfe 2001; Mandáková et al. 2010).

The prevalence and relative contributions of auto- and allopolyploidy to plant speciation vary. Estimates show that in natural populations, the prevalence of allopolyploidy is higher than autopolyploidy (Levin 2002; Ramsey and Schemske 1998). However, autopolyploidy might be underestimated because it is quite difficult to identify in natural settings (Tate et al. 2005a; Soltis et al. 2007). The apparent prevalence of allopolyploids could be because of heterosis conferred by their hybrid origin that gives them an edge over autopolyploids (Stebbins 1950; Grant 1981b). Also, autopolyploids are more prone to chromosome pairing problems in meiosis (Grant 1981b) and therefore might not persist over longer evolutionary timeframes.

1.3. Formation of Polyploids

Gametic non-reduction and somatic doubling, to some extent, are the main means of polyploid formation (Lewis 1980; Ramsey and Schemske 1998; Levin 2002; Coyne and Orr 2004). In meiotic nuclear restitution, commonly known as gametic non-reduction, alterations in the meiotic program or cellular defects in meiosis I (MI) or meiosis II (MII) may switch the meiotic cell division into a mitotic-like process, generating diploid or unreduced gametes (Brownfield and Köhler 2010), thus leading to progeny with an increased chromosome number. This mechanism can be subdivided into two types: bi- and unilateral sexual polyploidization (Ramsey and Schemske 1998; Comai 2005b; De Storme and Geelen 2013). In the former, fusion between two gametes having a fully somatic complement of chromosomes can result in a tetraploid individual, which, depending on favorable conditions, can initiate the de novo establishment of a tetraploid lineage. However, in unilateral sexual polyploidization, the fusion of a normal haploid gamete with an unreduced gamete leads to the formation of a triploid embryo. Although triploid seeds are not viable due to imbalanced parental genome input in the endosperm, this triploid block is often absent or incomplete, permitting triploid plant formation (Simioni et al. 2006; Köhler et al. 2010). Through meiotic cell division or random segregation, these triploids generate aneuploid and/or euploid gametes, which may result in the establishment of stable polyploid populations over time (Henry et al. 2005). This process is commonly termed the triploid bridge hypothesis (Ramsey and Schemske 1998) and has been proposed to play a role in the evolution of polyploidization events. These pathways might be common as in almost every plant species, the first or second meiotic division results in variable but small amounts of unreduced gametes (Mok and Peloquin 1975; Carputo et al. 2000). Cytological studies from various plant species suggest a plethora of cellular defects can lead to this phenomenon. Three main divisions of these

defects are: (1) meiotic spindle dynamic alterations; (2) defects in the formation of the meiotic cell plate; (3) omission of meiosis I or II, (reviewed in Brownfield and Köhler 2010; De Storme and Geelen 2013; Bretagnolle and Thompson 1995). It has been estimated that non-hybrid flowering plants produce 0.56% diploid gametes (diplogametes) (Ramsey and Schemske 1998). Environmental stresses like nutrient shortage, frost, herbivory, and wounding enhance the frequency of diplogamete production (Mayrose et al. 2011; Pécrix et al. 2011). Strikingly, unreduced gamete production is 50-fold higher in hybrids compared to non-hybrid systems (Zhang et al. 2010b). Additionally, the production of diplogametes is heritable and their production is controlled by a small number of genes (Zhang et al. 2010b; d'Erfurth et al. 2008). Although the underlying molecular phenomenon of unreduced gamete formation is known, much is yet to be revealed (Brownfield and Köhler 2010).

The other main avenue of polyploid formation is via somatic doubling. Mixoploid chimeras are produced by somatic doubling in meristem tissue of either adult or juvenile sporophytes (Ahloowalia and Garber 1961; Hiesey et al. 1971; Jorgensen 1928; Newton and Pellew 1929; Skalińska 1947). Fertile tetraploid shoots of otherwise sterile diploid F_1s, of *Primula floribunda* × *P. verticellata*, generated *Primula kewensis*, one of the first described allopolyploids (Newton and Darlington 1929). Likewise, a tetraploid shoot was observed in wounded ("decapitated") tomato plants (Jorgensen 1928) and on a diploid F_1 hybrid between *Mimulus nelsoni* and *M. lewisii* (Hiesey et al. 1971). Somatic polyploidy is commonly observed in numerous non-meristematic plant tissues (D'Amato 1952, 1964), including normal diploid *Vicia faba*, which possesses octoploid and tetraploid cells in pith and cortex of stem (Coleman 1950). Such polyploid cells are a source of new polyploid shoots (D'Amato 1952, 1964; Lewis 1980). Complete polyploid sporophytes can also be generated through the process of somatic doubling occurring

either in a young embryo or zygote. How frequently somatic doubling occurs naturally and the effects of interspecific hybridization on its frequency are still not well understood (Lewis 1980; Ramsey and Schemske 1998).

Synthetic polyploids, to imitate the established natural polyploids, are mostly produced by somatic doubling, although this pathway is less prevalent in natural polyploids (Tayalé and Parisod 2013). The generation of synthetic polyploids can be done by the use of colchicine, trifluralin, oryzalin or amiprophos-methyl, i.e., anti-mitotic substances that block the cell cycle (Jaskani et al. 2005; Semeniuk and Arisumi 1968; Madon et al. 2005; Amiri et al. 2010; Rodrigues et al. 2011; Tamayo-Ordóñez et al. 2016). Colchicine is the most commonly used means that inhibits the segregation of chromosomes in mitosis through the depolymerization of the microtubular cytoskeleton in early stages of metaphase, consequently leading to polyploidization of the cells. In higher concentrations, in a later stage, it induces polymerization of new tubulin-containing structures in c-metaphase cells, allowing the reconstitution of 4C nuclei and their progression into the cell cycle (Caperta et al. 2006).

1.4. Genomic and Genetic Consequences of Polyploidy

Interspecific hybridization and polyploidy have played a pivotal role in modifying genomes of flowering plants (Soltis et al. 2015). Genomic studies of modern and ancient polyploids demonstrate immediate and long term consequences of polyploidy at the genomic and transcriptomic level that result from merging two genomes (Wendel et al. 2018a). In the short term, some of the immediate consequences include duplicated gene loss, divergence in molecular evolutionary rates, homoeologous exchange of reciprocal or non-reciprocal nature, gene silencing and inter-subgenomic spread of TEs (Figure 1.1) (Parisod et al. 2010a; Senerchia et al. 2015). Longer-term consequences of

polyploidy include structural rearrangements in the polyploid genome, the emergence of novelty in genome function (neofunctionalization) or the sharing of gene function between duplicates (sub-functionalization) (Yoo et al. 2014a; Renny-Byfield et al. 2014). Over time, at a structural level, loss of duplicated genes and repetitive sequences, a decrease in the number of chromosomes and extensive chromosomal rearrangements are observed frequently, which may lead to genome downsizing in polyploids (Freeling et al.

Figure 1.1: Variation in allopolyploids at the genome level including changes in the karyotype, patterns of methylation and transposable element distribution (from Soltis et al. 2014a).

2012; Woodhouse et al. 2014; Leitch and Leitch 2008). All these genomic alterations can occur collectively over time, ultimately resulting in polyploid systems that behave like diploids cytogenetically (Wendel 2015). The diverse nature of polyploid systems affects both the speed and the extent to which these variations occur (reviewed in Soltis et al. 2016). Moreover, the diploidization process can also occur with respect to loss/retention

of certain types of genes and gene losses/retention from a particular genome (Wendel 2015) (discussed in sections 1.4.2 and 1.5.3). These various aspects of genome downsizing or diploidization suggest that polyploidy, to some extent, is a reversible process (Leitch and Leitch 2008). In polyploid genomes, the occurrence of similar sequences derived from parents enhances the likelihood of structural variants. The following genomic effects have been extensively studied and observed in multiple polyploid species.

1.4.1. Chromosomal Rearrangements and Homeologous Exchanges

Chromosomal variations in polyploids are mostly caused by abnormal chromosomal segregation at meiosis due to multivalent formation during early generations (Comai 2005b). Thus, the segregation of chromosomes and meiotic/mitotic stability are major challenges for cells after polyploidization (reviewed by Comai 2005b). In allopolyploids, the phenomenon of disomic inheritance, where homologous chromosomes pair and recombine during meiosis, is commonly observed (Griffiths et al. 2006). However, some polyploids have been shown to deviate from an expected disomic pattern and display various chromosomal rearrangements including intergenomic translocations, aneuploidy, and homeologous (duplicated chromosomes in polyploids derived from different parent species) recombination (Soltis and Soltis 2000). For instance, Chester et al. (2012b) reported that 69% and 76% of *Tragopogon miscellus* populations showed aneuploidy of one or more chromosomes and intergenomic translocations, respectively, with no apparent population having a fixed karyotype. Importantly, the populations showing aneuploidy were observed to be compensatory aneuploids (Figure 1.1) as they still had the stable chromosome number ($2n$=24), because aneuploidy occurred mostly through reciprocal monosomy-trisomy of homeologous

chromosomes (1:3 copies) or nullisomy-tetrasomy (0:4 copies) (Chester et al. 2012b). *Tragopogon mirus* showed similar results with a few exceptions in the frequency of compensated aneuploidy (Chester et al. 2015). These findings confirm that after WGD the prolonged and significant instability in chromosomes is a common effect in natural populations. Moreover, *Nicotiana* allopolyploids and synthetic lines of *Brassica napus* also showed similar intergenomic translocations and aneuploidy (Xiong et al. 2011; Gaeta et al. 2007). However, plant allopolyploids such as cotton and *Spartina anglica* were observed to display relatively few exchanges between homeologous chromosomes which might be due to divergence between parental genomes, thus led to stable allopolyploid genomes (Liu et al. 2001; Ainouche et al. 2012; Ainouche et al. 2004). These findings suggest that in angiosperms, different plant species react differently to polyploidy.

1.4.2. Gene Loss

As duplicated genes are located at all loci of allopolyploids, these extra loci can be lost, maintained and/or modified (Figure 1.2). Loss may consist of either homeologous gene conversions or physical loss of DNA (Wang and Paterson 2011). It has been reported both in synthetic and natural allopolyploids that gene losses of one duplicated copy have occurred over short time scales (Kashkush et al. 2002a; Nie et al. 2008; Anssour et al. 2009; Buggs et al. 2010a; Koh et al. 2010). Nie et al. (2008) identified immediate gene loss following polyploidization in a S_1 synthetic hexaploid *Triticum*. On the contrary, gene losses in other groups occurred in later generations after WGD (Song et al. 1995; Buggs et al. 2009b). The frequency of gene losses can be high in young polyploids, as evident in *Tragopogon miscellus*, which showed high variation in homeolog losses among populations and individuals (Tate et al. 2009b; Buggs et al. 2012a). Moreover, in multiple origins of both *T. mirus* and *T. miscellus*, recurrent patterns

of homeolog retention and loss were reported (Tate et al. 2006a, 2009; Koh et al. 2010; Buggs et al. 2012a).

Figure 1.2: Variation in allopolyploids at the genetic level involving homeolog loss, retention or conversion across multiple lineages (from Soltis et al. 2014a).

Over longer evolutionary time, gene function plays major roles in governing the fate of duplicated genes, whether they are retained in duplicate or returned to single copy (Paterson et al. 2006; Barker et al. 2008; Severin et al. 2011; De Smet et al. 2013a; Wendel 2015). Genes that encode subunits of multi-subunit proteins (such as transcription factor complexes, ribosomes and proteosomes) or related proteins are more likely to be retained (Van de Peer et al. 2017a; Freeling et al. 2015; Conant et al. 2014). Whereas, genes related to DNA metabolism, RNA binding proteins and nuclease activity are singleton genes and are anticipated to work alone (Freeling 2009). Even though the genomic events associated

with neopolyploids may seem random, over longer periods, patterns of duplicated vs single gene retention eventually appear.

Bias in gene loss from particular parental genomes has also been documented in studies investigating gene duplications in *A. thaliana* (Thomas et al. 2006), *B. rapa* (Cheng et al. 2012) and *Z. mays* (Schnable et al. 2011). In *B. rapa*, one of the three sub-genomes encountered considerably fewer gene losses than the others (Tang et al. 2012; Wang et al. 2011). A similar pattern is also mirrored at the expression level, where a gene producing more transcripts tends to be retained (Schnable et al. 2011) and located on the dominant sub-genome (Woodhouse et al. 2014). This aspect of expression dominance is linked to differences in the density of transposable elements between parental genomes, such that the dominant genome with less TEs would reflect less methylation and higher gene expression levels compared to the co-resident genome.

The divergence between progenitor genomes of the allopolyploid may impact the differential fate of homeologs. For instance, some homeologs may already be well adapted to particular ecological conditions or may contribute more to the total expression. Also, because of the inherited initial differences in local genome context between two progenitor genomes of an allopolyploid, selection and mutation might act differentially on homeologs (Grover et al. 2017). For example, differences in epigenetic modifications, *cis*/*trans* regulation, the recombinational landscape, gene density and gene activity between parents could lead to differential regulation of homeologs, with subsequent unequal effects of diversity-altering phenomena (He et al. 2010; Guo et al. 2008; Grover et al. 2017). Lastly, gene conversion may alter homeolog interactions in the nucleus of an allopolyploid via creating chimeric and novel alleles but minimizing intragenomic nucleotide diversity. Gene conversion (non-reciprocal recombination between alleles or paralogous loci) is apparently a common phenomenon in species of allopolyploids,

including cotton, *Nicotiana rustica*, *Brassica napus*, *Haloferax volcanii*, and *Coffea arabica* (Matyasek et al. 2003; Udall et al. 2005; Gaeta and Pires 2010; Salmon et al. 2010; Lange et al. 2011; Page et al. 2016; Flagel and Wendel 2010; Lashermes et al. 2016). Certainly, stability in fungal and plant genomes following polyploidization or hybridization can be maintained by gene conversion (McGrath et al. 2014; Chalhoub et al. 2014; Sriswasdi et al. 2016). Qiao et al. (2019) reported dynamic changes in the rate of gene conversion, which was high immediately after polyploidization and decreased overtime, reflecting the 1 My (million years) of cotton evolution (Guo et al. 2014). Collectively, all these different avenues of gene loss contribute to the diversity of polyploid genomes post-formation.

1.5. Transcriptomic Consequences of Polyploidy

The transcriptome is the sum of all the RNA transcripts of an organism. Following WGD, polyploids go through changes in gene expression generated by massive genomic restructuring (chromosomal translocations, gene losses, homeologous exchanges) and epigenetic mechanisms (cis/trans effects, histone modifications, DNA and transposons methylation), with subsequent effects on the metabolome and proteome (Chen 2007a). In allopolyploids, transcriptome shock occurring via the combination of diverged genomes also contributes to the rapid and substantial variation in the expression levels between homeologs, though to varying extents (Hegarty et al. 2006; Jackson and Chen 2010). Interspecific hybridization, rather than chromosome multiplication itself, has a major impact on the reprogramming of the allopolyploid transcriptome (Chen 2007a; Osborn et al. 2003b). Thus, moderate transcriptomic modifications triggered in autopolyploids are not as apparent as in allopolyploids, due to their intraspecific origin (Comai 2005b).

Previous studies have revealed four possible fates of duplicated genes in allopolyploids: both homeologs are functional, one parental copy is silenced while other maintains the original function, both copies diverge and partition their ancestral function or one copy may gain a novel function (Prince and Pickett 2002; Doyle et al. 2008; Conant and Wolfe 2008; Edger and Pires 2009; Lynch and Conery 2000; Blanc and Wolfe 2004; Roulin et al. 2013). Duplicated genes may not conserve their parental legacy and their expression levels can show divergence from parental additivity (Buggs et al. 2014). In allopolyploids of wheat (Pumphrey et al. 2009; Qi et al. 2012), cotton (Flagel et al. 2008), *Tragopogon* (Buggs et al. 2011a), *Arabidopsis* (Shi et al. 2012a; Wang et al. 2006b) and *Senecio* (Hegarty et al. 2006) non-additive expression patterns have been reported, and so exhibit expression phenotypes that are not additive to those of the parental taxa. Such dynamic expression discrepancies in allopolyploids generate genetic and phenotypic novelty, which promotes polyploid evolution. In many allopolyploids, transcriptomic responses to interspecific hybridization followed by genomic doubling have been revealed, including a variety of phenomena, such as genome dominance, restoring of co-expression networks, homeolog expression bias and expression level dominance (Figure 1.3) (Hu and Wendel 2019a; Grover et al. 2012a; Soltis et al. 2012; Wendel et al. 2012). These scenarios are elaborated upon in the following sections.

1.5.1. Homeolog Expression Bias

In an allotetraploid, when expression of one homeolog (duplicated genes) is favored over the other, this phenomenon is called homeolog expression bias (Figure 1.3). The outcome differs among tissues and is expected to be based on the expression levels inherited from the progenitor diploids (Flagel et al. 2008; Chaudhary et al. 2009; Buggs et al. 2011a; Dong and Adams 2011; Combes et al. 2011). This definition of expression

bias has an obvious evolutionary dimension in the sense that it involves expression level comparison among polyploid and progenitor or model progenitor genomes. In cases where parental levels of expression cannot be determined, biased expression is relaxed relative to ancestral states by assuming 1:1 parental expression (Schnable et al. 2011). Moreover, when progenitor expression levels are unequal, bias is often determined by

Figure 1.3: Variation in allopolyploids at the transcriptome level involve differential expression of homeologs resulting in homeolog expression bias, expression level dominance or tissue specific expression patterns. The cis/ trans regulatory differences and their interactions may cause such expression level changes (Soltis et al. 2014a).

comparing the expression of each homeolog to a mid-parent expression value or to the expression level of the parents (Rapp et al. 2009a; Chagué et al. 2010; Flagel and Wendel 2010). The causes of biased expression may be due to epigenetic mechanisms (discussed in section 1.6.1), including cis/trans regulatory differences between merged genomes (Figure 1.3) (Shi et al. 2012a; Bell et al. 2013; Xu et al. 2014). Homeolog expression

bias has been reported in many allopolyploid systems, including *Arabidopsis* (Wang et al. 2004; Chang et al. 2010), *Brassica* (Gaeta et al. 2007; Auger et al. 2009), *Gossypium* (Adams et al. 2003b; Flagel et al. 2008; Hovav et al. 2008), *Tragopogon* (Buggs et al. 2010a, b; Koh et al. 2010), *Spartina* (Chelaifa et al. 2010), *Triticum* (Mochida et al. 2004; Bottley et al. 2006) and others.

Across diverse species, synthetic and natural allopolyploids behave differently with regard to expression bias. For instance, an RNA sequencing study in cotton allopolyploids showed that one sub-genome expressed more genes and the nature of the sub-genome varied in synthetic versus natural allopolyploids (Hu et al. 2013). In comparison to diploids, 48% and 59-62% of genes showed differential expression in synthetic and natural polyploids, respectively (Hu et al. 2013). Among 25,000 cotton genes encompassing single nucleotide polymorphisms distinguishing homeolog origins, only 0.71–0.75% genes have no visible expression in synthetic polyploids. Since the allopolyploidization event in cotton occurred 1–2 My ago, it was concluded that there may not have been enough time for gene loss or expression of pseudogenized copies are still detectable (Wendel and Cronn 2003a). Homeolog-specific patterns showed immediate bias in expression of one homeolog over the other in *Tragopogon miscellus* following hybridization (Buggs et al. 2011a), with a minimum influence of polyploidy per se (Buggs et al. 2014). Across synthetic lines and natural populations, evolutionary patterns are significantly repeated, with *T. mirus* exhibiting similar outcomes (Tate et al. 2009b; Buggs et al. 2009b).

Genome-wide investigations of expression differences are challenging for most organisms, because of the scarcity of genome sequence data for most species. In these cases, small scale experiments to investigate expression bias in non-model allopolyploid systems have been conducted. For example, only 13 genes in *Coffea arabica* (Combes et

al. 2011), 30 in *Tragopogon mirus* (Koh et al. 2010) and 144 in *T. miscellus* were analysed (Buggs et al. 2011a). With only a few studies examining duplicate gene expression, our understanding of how homeologs are regulated in allopolyploids is limited. Due to the advancement of RNASeq technology, studies of homeolog expression bias in non-model allopolyploids can be carried out without prior information regarding complete parental genome sequences.

1.5.2. Expression Level Dominance

Expression level dominance (originally termed as genomic dominance) occurs when the total expression of a homeologous gene pair turns out to be more similar to one parent irrespective of the expression level of the gene in the parent in question (Figure 1.3) (Grover et al. 2012a; Li et al. 2014; Rapp et al. 2009a). Both expression level dominance and homeolog expression bias may be balanced among loci in a genome with regard to diploid progenitors or they may favor one parental genome over the other (described as unbalanced) (Grover et al. 2012a). Initially, expression level dominance was observed in cotton (Rapp et al. 2009a; Yoo et al. 2013a) and now has been demonstrated in other plant species including wild allopolyploids (Chelaifa et al. 2010), coffee (Bardil et al. 2011), and wheat (Qi et al. 2012). Expression level dominance in coffee plants was only observable when grown at higher (Bardil et al. 2011). Such observations add a fascinating environmental dimension to the stability of gene expression in polyploids. Like coffee and cotton, ~21-22% of transcripts examined in leaf tissue of *Tragopogon* allopolyploids displayed unbalanced expression level dominance (toward the maternal parent) (Boatwright et al. 2018a). In cotton and coffee, both of which are older allopolyploids compared to *Tragopogon* polyploids, the modulation and selection of regulatory networks over time (Flagel and Wendel 2010) may lead to an

increase in non-additive expression compared to recent allopolyploids that exhibit a lower degree of non-additive expression. Similarly, Flagel and Wendel (2010) described a sequential dimension to this phenomenon in a study of five natural allopolyploid species, showing that although recent allopolyploids display maximum level of bias in genomic dominance, it dissipates over evolutionary time even though the overall extent remains comparatively high.

A connection between homeolog expression bias and expression-level dominance was shown by three studies conducted using RNA-Seq; two studies indicated no link between them (Combes et al. 2013; Rambani et al. 2014), while the third study showed that either down- or upregulation of the non-dominant parental copy might lead to expression level dominance (Yoo et al. 2013a). These findings illustrate the action of various molecular mechanisms comprising silencing of the non-dominant homeolog, or trans-effects overpowering cis-regulatory differences. Despite the limited knowledge of underpinning mechanisms, homeolog expression bias and expression-level dominance proved to be a source of novelty and diversity in polyploids relative to their diploid parents, hence they can have significant roles in polyploid evolution.

1.5.3. Sub-genome/Genome Dominance

In allopolyploids that show biased fractionation, one of the prominent whole genome features is sub-genome dominance, where genes of one sub-genome seem to be expressed at a higher rate, depending on their maternal or paternal origin, as compared to the genes of second sub-genome (reviewed in Yoo et al. 2014a). Feldman et al. (2012) reported this interesting scenario in wheat allopolyploids in which the A genome was shown to control morphological traits, including grain shape, glumes with keels, free caryopsis, plant habitus and inflorescence structure. In addition, this genome also harbors

many domestication genes, e.g., free-threshing on long arm of chormosome 5A (5AL) (Sears 1954), QTLs for kernel size (Elias et al. 1996), the genes for non-brittle spike on short arm of chromosome 3A (3AS) (Nalam et al. 2006). Hence, the A genome likely has a complete suite of important genes, thereby regulating differential sub-genome control of morphological traits, whereas the B genome in wheat allopolyploids confers tolerance to biotic and abiotic stresses, thus playing a role in ecological adaptation of populations (Peng et al. 2003b; Peng et al. 2003a; Fahima et al. 2006). Genome dominance requires some generations to be established since it is a characteristic feature mostly found in ancient polyploids instead of synthetic ones (Woodhouse et al. 2014). A real-time study involving diploid parents, a hybrid and a recent allotetraploid is the best way to understand the origin of sub-genome dominance. While studying monkey flower tetraploids, Edger et al. (2017) confirmed that sub-genome expression dominance occurred in a wide hybrid (produce via intergeneric hybridization) immediately following hybridization, and substantially increased over generations. This was evidenced by low methylation levels of TEs and nearby genes in the wide hybrid, followed by intermediate levels in the synthetic allotetraploid. However, natural allopolyploid (< 140 years) showed highest differential levels of methylation between dominant and submissive sub-genomes (Edger et al. 2017). Such differences in methylation levels between hybrid, synthetic- and natural allotetraploids reflected a progressive increase in expression bias patterns. Moreover, they concluded that after millions of years, in allotetraploids derived from a wide hybrid, the submissive genome expressing fewer genes will be inclined toward biased gene fractionation compared to its homeolog. Hence, while sub-genome dominance occurs instantaneously in a hybrid, it requires evolutionary time to increase.

Even though it is difficult to measure sub-genome differences in terms of fractionation until a few million years have passed, variation at the gene expression level

in recent allopolyploids can be measured. Several cases of sub-genome dominance have been demonstrated by various allopolyploids including wheat (Li et al. 2014), cotton (Renny-Byfield et al. 2015) and *Tragopogon mirus* (Buggs et al. 2010b). Likewise, mesopolyploid (*B. rapa* and maize) sub-genomes have been differentially fractionated, and the sub-genome with more genes tended to display comparatively higher expression levels of mRNA (Schnable et al. 2011; Cheng et al. 2012). In contrast to the above-mentioned studies, sub-genomes of pumpkin allotetraploids (> 3My old) showed no biased fractionation or sub-genome expression dominance and interestingly, they are chromosomally intact, even though deletions have taken place (Sun et al. 2017). It could be because of no noteworthy differences in TE distributions in cucurbit diploid progenitor species (reviewed in Wendel et al. 2018a).

Differential density of TEs and their methylation status in original diploid progenitors of an allopolyploid as well as chromatin modifications are among those mechanisms that have been proposed to mediate inter-genomic suppression of gene function (Feldman et al. 2012), thus leading to the origin of genome dominance (Renny-Byfield et al. 2015, 2017; Edger et al. 2017; Steige and Slotte 2016; Cheng et al. 2016; Wendel et al. 2018a; Woodhouse et al. 2014; Pophaly and Tellier 2015). Previous studies have mainly investigated the diverse expression patterns of duplicated genes in old allopolyploid species. Young *Tragopogon miscellus* allopolyploid could prove potentially valuable to study sub-genome dominance as its history of origin is tractable (90-100 years old) and also the parental diploid species are known which still coexist with their allopolyploid derivatives.

1.5.4. Developmental and Tissue-specific Regulation of Gene Expression

In various plant allopolyploids, the evolutionary outcome of homeologs has been

reported in numerous studies investigating homeolog-specific expression analyses (Adams 2007; Hughes et al. 2014; Takahagi et al. 2018b). A study examining genome-wide expression changes in *Arabidopsis* allotetraploids discovered that during flower development, silenced rRNA genes in leaves get activated in flower organs including sepals, pollen, petals and anthers, (Wang et al. 2006b). In *B. napus*, multiple homeologs encoding ribosomal proteins show tissue-specific expression, revealed through transcriptome analysis (Whittle and Krochko 2009). Furthermore, in cotton, analysis of relative expression levels of homeologous and allelic genes revealed that genes showing sub-functionalization expressed in reproductive tissues and non-functionalized alleles are the derivatives of the A-genome, suggesting most likely genome-of-origin bias for neo-functionalization (Chaudhary et al. 2009). An alternative study conducted on cotton confirmed that organ-specific gene expression is exhibited by many homeologous genes that in allotetraploids display unequal expression (Adams et al. 2003b). One homeolog of the locus, *AdhA*, was interestingly silenced in one organ and other homeolog in another organ. This reciprocal silencing mechanism occurs in both natural and synthetic allotetraploids of cotton and is dependent on genotype (Adams et al. 2004), thus indicating stable maintenance of gene duplicates during evolution and their rapid sub-functionalization (Huerta-Cepas et al. 2011). Similarly, studies based on recent *Tragopogon* polyploids reveal that immediately following WGD, sub-functionalization may begin to happen (Buggs et al. 2011a). In allopolyploid species, the evolutionary changes in gene regulatory networks caused by differential expression of diverged homeologs are believed to mediate responses to environmental cues and developmental programs (Chen and Ni 2006a).

1.6 Mechanisms for Regulation of Non-additive Gene Expression

In allopolyploids, numerous controlling factors acting either discretely or collectively may lead to non-additive gene expression. One of these factors is the expression differences inherited from parental diploid species that result in differential expression of homeologs in allopolyploids (Figure 1.3) (Buggs et al. 2014). This concept was developed decades ago (Gottlieb 2003) but still has implications for elucidation of gene expression data, i.e., without carefully examining the progenitor expression patterns, differential homeolog expression should not be described as post-polyploidization shifts in expression and it may not necessarily show a departure from parental patterns. Therefore, a combination of parental legacy and other influences, as described below, may result in these variable expression patterns in polyploids.

1.6.1. Epigenetic Regulation of Duplicated Genes

In allopolyploids, reprogramming of developmental patterns and gene expression is induced by polyploidy through epigenetic modifications of homeologous chromosomes (Song and Chen 2015). This results into innovative dosage-dependent and independent expression variations at the genome level (Shi et al. 2015). Studies on both natural and synthetic polyploids have demonstrated that such changes at the functional level are reproducible, to a large extent, and happen quickly (e.g., *Brassica, Tragopogon* or *Triticum*) (reviewed in Tayalé and Parisod 2013; Soltis et al. 2012; Grover et al. 2012a). Contrary to canonical genetic mechanisms, epigenetic marks control the access to genetic information more than the genetic sequence alteration itself (reviewed in Galindo-González et al. 2018). RNA-interference (RNAi), DNA methylation and histone modifications are three main epigenetic mechanisms described in plants (Pikaard and

Scheid 2014). In addition to these, non-additive expression patterns (e.g., homeolog expression bias) can also be induced by modulation of trans- and/or cis-regulatory elements. It has been described that different patterns of homeolog suppression and activation can result from the combination of two divergent regulatory systems (reviewed in Yoo et al. 2014a). Low sequence divergence presented by two parental genomes, such as those found in cotton [20% between 89,588 (A-genome) and 65,542 (D-genome) contigs (Grover et al. 2007; Guan et al. 2014)], *Tragopogon* [1.45% in 50,766 coding nucleotides (reviewed in Buggs et al. 2014; Boatwright et al. 2018a)] and coffee [1.35% in 60,000 coding nucleotides (Cenci et al. 2012)] may, in resulting allopolyploids, assist cross-talks among parental trans-regulatory elements. Hence, it is likely to infer that intertwined mechanisms of regulatory elements from both parents regulate homeologs. Despite their significant role in duplicated gene regulation, only a few polyploids have been studied for homeolog expression bias and cis/trans regulatory factors [e.g., cotton (Chaudhary et al. 2009; Yoo et al. 2013a), *Arabidopsis suecica* (Shi et al. 2012a), and coffee (Combes et al. 2013)].

DNA methylation, one of the epigenetic mechanisms, represses the transcription of genes by inhibiting the binding of transcription factors (TFs) to the promoter through methylation of the coding or promoter regions of the genes, thus silencing the duplicated genes (Salmon et al. 2005; Salmon and Ainouche 2010). In plants, DNA methylation occurs in CHH, CG and CHG (H = A, T or C) contexts by diverse pathways. Molecular processes such as those involved in plant and animal development, including gene imprinting, transposon silencing and virus defense, are affected by DNA methylation (Law and Jacobsen 2010; Richards 1997). Following both allo- and autopolyploidization, DNA methylation shows non-additive patterns (Salmon et al. 2005; Kraitshtein et al. 2010; Lavania et al. 2012; Zhao et al. 2011) and is anticipated to play a role in homeolog

silencing in situations where both copies have counterproductive activities (Chen and Tian 2007). For instance, monomeric subunits encoded by homeologous genes assemble to form dimeric or polymeric proteins (Phillips et al. 1995). In allopolyploids, fitness can be increased by preventing the formation of heterodimers, through silencing of one homeolog of such subunit encoding genes. It has also been stated that DNA methylation may act in a concerted way with histone modifications to regulate gene expression. A study conducted on *Arabidopsis* determined that during hybridization, DNA methylation acts upon a different subclass of heterochromatic genes than those affected by histone acetylation (Chen and Tian 2007). These authors proposed that in early generations of polyploids, histone acetylation leads to remodeling of chromatin structure of these loci, thus inhibiting transcription immediately. Subsequently, silencing is done in a stable manner through DNA and histone methylation.

In polyploids, regulation of gene expression via cytosine methylation has received great attention in recent years. It has been observed in wheat that cytosine methylation acts as an immediate response to whole genome duplication (Shaked et al. 2001). Verhoeven et al. (2010) demonstrated that various ecological stresses induce heritable methylation modifications between genetically identical triploid dandelions. Moreover, a study conducted on *Oryza sativa* to estimate the methylation level of coding regions of different gene categories in five plant tissues showed that low and high levels of methylation are linked with low and high levels of gene expression, respectively (Wang et al. 2013). This suggests that duplicate gene expression and survivorship is related to genome-wide methylation patterns (Wang et al. 2013). Likewise, in allopolyploids of *Spartina anglica* (Parisod et al. 2009; Salmon et al. 2005), *Tragopogon* (Sehrish et al. 2014), *B. napus* (Lukens et al. 2006), and *Arabidopsis* (Madlung et al. 2002), extensive DNA methylation changes have also been reported.

Similar to DNA methylation, histone modifications are expected to play substantial roles in transcriptome shock resulting from whole genome duplications (Ha et al. 2011). Eukaryotic DNA is wrapped into chromatin, a highly compacted structure, through interacting with histones (Luger et al. 1997). Covalent modifications of core histones (H2A, H2B, H3 and H4) predominantly through acetylation and methylation, have been reported to have various effects on gene expression (Chen and Tian 2007; Liu et al. 2010). Histone marks such as H3K4me3 and H3K9ac, are generally located on euchromatin, thus linked to active transcription, while H3K27me3 and H3K9me2 modifications, found on heterochromatin, are associated with gene repression (Jenuwein and Allis 2001; Liu et al. 2010; Li et al. 2007).

The genomic shock linked with polyploidization and interspecific hybridization may be responsible for causing modifications in chromatin structure and subsequently gene expression (McClintock 1983; Chen 2007a; Comai 2005b). The role of chromatin modifications in silencing or activating the homeologs in allopolyploids has been reported in various studies (Kashkush et al. 2002b; Lee and Chen 2001; Madlung et al. 2002; Scheid et al. 2003; Wang et al. 2006a; Wang et al. 2004). For example, Zhang et al. (2017) demonstrated that in a wheat hexaploid, differential expression of *TaGS2* homeologs is regulated by H3K4me3 in different developmental stages and organs. Likewise, in allotetraploids, high and low levels of H3K4 methylation and H3K9 acetylation in promoters causes up-regulation (*TOC1* and *GI*) and repression (*CCA1* and *LHY*) of central circadian clock genes, respectively (Ni et al. 2009). Moreover, in *Arabidopsis* allopolyploids, increased levels of H3K4 dimethylation and H3K9 acetylation in promoters regulate up-regulation of *FLC* (FLOWERING LOCUS C) that prevents early flowering (Wang et al. 2006b). Thus, late flowering was observed in synthetic allotetraploids compared to their progenitors, while flowering in natural *A. suecica*

allotetraploids starts even later than synthetic lines (Wang et al. 2006b). The above-mentioned data suggest that investigating the role of histone modifications, in regulating the homeolog expression by profiling progenitors and allopolyploids might be an interesting question. This will help unravel the interfaces that link epigenetics, functional pleiotropy and cellular phenotype.

1.6.2. Transposable Elements and siRNAs

Transposable elements (TEs) are prevalent in eukaryotic genomes and in order to persist, their sequences undergo "selfish" replication (Joly-Lopez and Bureau 2014; Doolittle and Sapienza 1980), thus, acting as an important driver of evolution and contributing to genome plasticity (Gifford et al. 2013; Oliver et al. 2013; Friedli and Trono 2015). It is also widely evident that TEs control many plant traits through having cis-regulatory effects on host genes (Chuong et al. 2017). Although most TEs are silent in their host plants, certain environmental stresses and genomic shock resulting from interspecific hybridization, can lead to genomic instability and activation of merged TEs (McClintock 1983) and subsequently, these TEs may modulate gene expression, specifically non-additive expression patterns. Polyploidy immediately activates specific TEs (Parisod et al. 2010a). For example, as reported in wheat allopolyploids, transcriptional activation of retrotransposons (Kashkush et al. 2002b) suggests that in allopolyploids gene expression modulation might be triggered by epigenetic alterations in TE expression (Parisod et al. 2009). So far, there is little evidence for gene expression changes by TE activation or repression in polyploids.

Various mechanisms regulating TE activity have been described. One of these mechanisms includes the role of repressive small interfering RNAs (siRNAs), that target and silence homologous inserted copies via DNA methylation (Bourc'his and Voinnet

2010). Small interfering RNAs (siRNAs; 21-24 nucleotides) are a class of small RNAs arising from double-stranded RNA (dsRNA) precursors via the activity of Dicer-like proteins (DCL2, DCL3, and DCL4) (Shamandi et al. 2015). The diversity of small RNAs and their ability to function in trans, and to communicate their silencing effects across members of gene or TE families, have been emphasized by various reviews (Axtell 2013; Fei et al. 2013). The siRNA silenced TEs that are located near the functional transcriptional units may have the tendency to silence the expression of adjacent genes by position effects through a mechanism postulated as small RNA (smRNA)- directed DNA methylation (RdDM) (Hollister et al. 2011; Zhang et al. 2015b; Hollister and Gaut 2009). Because all TEs in plants are not able to initiate spreading (Eichten et al. 2012), the underlying mechanisms responsible for 'spreading' of silenced chromatin from transposons to nearby chromatin (Talbert and Henikoff 2006) are not yet understood in plants.

The concept regarding sub-genome dominance observed in ancient polyploids, is based mainly on a 'trade-off' hypothesis presented by Hollister and Gaut (2009). This hypothesis states that genomes experiencing the huge impact of transposons must make a trade-off between silencing the active TEs and their effects as silencers of expression of nearby genes. Thus, it has been proposed that biased expression of homeologs in allopolyploids is likely regulated by the methylation status of TEs depending on sub-genome location (Zhang et al. 2015b; Springer et al. 2016; Hollister and Gaut 2009). Also, the origin of sub-genome dominance is hypothesized to occur at the wide hybrid level, facilitated by the differences in TE loads, and/or methylation status of the original progenitor diploids (Renny-Byfield et al. 2017; Renny-Byfield et al. 2015; Edger et al. 2017; Steige and Slotte 2016; Cheng et al. 2016; Woodhouse et al. 2014; Pophaly and Tellier 2015; Wendel et al. 2018a). Therefore, the epigenetic regulation of parental

transposable elements, combined in one polyploid nucleus may consequently affect gene expression modulation in allopolyploids.

In ancient allopolyploids, the dominant sub-genome is characterized by the presence of both more genes and more transposons among those genes (Renny-Byfield et al. 2015; Murat et al. 2013; Cheng et al. 2016). A study conducted by Woodhouse et al. (2014) on sub-genome differences of *B. rapa* determined that lower expression of genes in the 'submissive' sub-genome was linked with both higher coverage of DNA upstream to transcriptional unit by 24nt-siRNA and transposons. However, in cases where genes with higher expression are located on the submissive sub-genome, the 'switchers', it was concluded that rather than sub-genome location, smRNA coverage was likely to be more associated with the expression of the gene (Cheng et al. 2016). Renny-Byfield et al. (2017) reported the same behavior pattern in maize as demonstrated in *B. rapa* by Woodhouse et al. (2014). Similarly, in *A. lyrata* and *A. thaliana*, siRNA targeted TEs are linked with decreased gene expression, the former species has two or three times more transposable elements than the latter (Hollister et al. 2011). Such massive adjustments associated with TE density and methylation occurring in allopolyploids are thought to become stabilized and return to near parental levels with successive generations. For instance, a genome-wide dramatic drop in CHH methylation was observed in monkeyflower hybrids but later on in polyploids, 'methylation repatterning' led to near parental methylation levels by readjustment in the TE densities (Edger et al. 2017).

It is becoming clear that conflicts between interspersed TEs and repressing siRNAs revealed by merging of divergent genomes, can lead to genome-wide reorganization (Parisod et al. 2010a). Such changes indeed lead to the rise of new stabilized states across the genome, but our understanding regarding the extent to which they are responsible for drastic changes in gene expression and phenotypes is limited. As

polyploidy instigates epigenetic and structural reorganization of both repetitive and coding fractions, young polyploids with known parental species, offer convenient models to address the question of how raw genetic material transforms into a significant evolutionary change.

1.6.3. Alternative Splicing in Regulating Gene Expression

In eukaryotes, alternative splicing (AS) acts as an important post-transcriptional regulator of gene expression and generates proteome and transcriptome diversity (Sultan et al. 2008; Syed et al. 2012; Kornblihtt et al. 2013). Alternative splicing produces numerous splicing variants by precise removal of introns and combinations of differential exons, from the same pre-mRNA (Wang et al. 2019). AS events are categorized into six major types, including alternative first exon (AltFE), intron retention (IR), alternative acceptor (AltA), alternative last exon (AltLE), exon skipping (ES) and alternative donor (AltD) (Sturgill et al. 2013; Wang et al. 2019). In plants, IR is observed to be the most common AS event with AS rates in genes containing introns ranging from ~30% to > 60% according to the available transcriptome data (Reddy et al. 2013; Sablok et al. 2011; Sablok et al. 2017). For example, it has been reported that 61% of multi-exonic genes in *Arabidopsis thaliana*, exhibited alternative splicing, including ~15.5% alternative 3' splice site, 7.5% alternative 5' splice site, ~40% intron retention, and 8% exon skipping/inclusion (Reddy et al. 2013). Transcriptomes of maize (Wang et al. 2016) and sorghum (Abdel-Ghany et al. 2016) showed similar results with intron retention being the most abundant mode of splicing, accounting for about 40%.

Gene expression can be influenced by alternative splicing on various levels: (a) by creating various forms of mRNA from a single gene, that further translates into

various types of protein isoforms; (b) by influencing mRNA stability through nonsense-mediated decay (NMD) pathway as indicated through studies conducted on *Arabidopsis* (Drechsel et al. 2013; Kalyna et al. 2011; Nyikó et al. 2009); and (c) by modulating mRNA stability and translation through miRNA regulation as suggested in studies on rice and *Arabidopsis* (Hirsch et al. 2006; Meng et al. 2013; Szarzynska et al. 2009; Yan et al. 2012; Yang et al. 2012). Consequently, in plants, AS plays significant roles in environmental fitness and developmental processes, including vernalization-mediated flowering (Rosloski et al. 2013; Marquardt et al. 2014), normal working of circadian clock (Seo et al. 2012; Filichkin et al. 2015), abiotic stress and biotic immune responses (Xu et al. 2012; Feng et al. 2015; Ling et al. 2015; Liu et al. 2018).

Alternative splicing provides an important means of driving evolution and adaptation in polyploids. Polyploidization could lead to dramatic and immediate changes in the activity and stoichiometry of splicing factors that further regulate AS profiling at the whole genome level (Syed et al. 2012). Considering an allotetraploid in which diploid parents contribute differing homeologs, various possible patterns of AS might arise in an allopolyploid. If no AS event is observed in diploid parents, then polyploidization may lead to new AS events in either one or both homeologs (Figure 1.4a). If AS events are detected in diploid parents, then the polyploid may retain, lose, or gain novel AS events in either one or both homeologs (Figure 1.4b).

AS changes affect many genes during the first few generations after allopolyploid formation. It has been reported that AS changes in homeologs are noticeably more common compared to homeolog silencing, which has been anticipated as affecting ~1–9% of homeologs (Wang et al. 2004; Kashkush et al. 2002a; Adams et al. 2004; Buggs et al. 2010a). However, AS changes examined in resynthesized *Brassica* allopolyploids have been shown to affect 26% of the genes (Zhou et al. 2011).

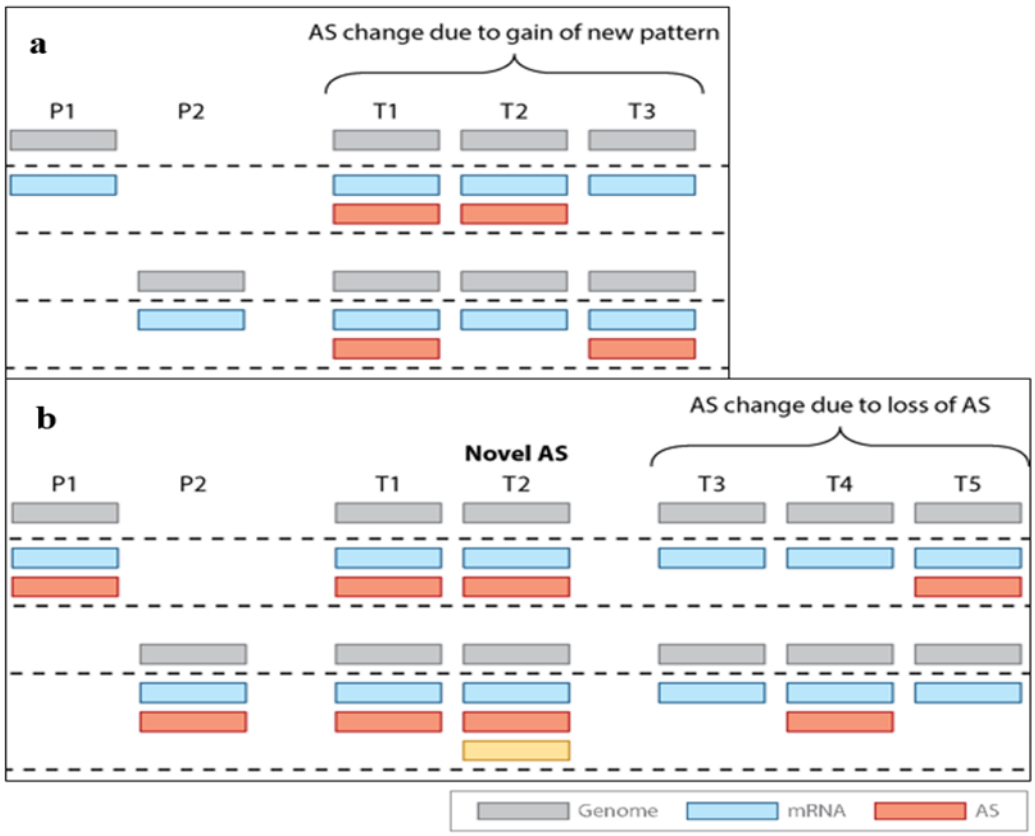


Figure 1.4: The outcome of alternative splicing in allopolyploids. (a) No alternative splicing (AS) events are present in the diploids at a given gene, but a new AS isoform is present post-polyploidization: gain of AS in homeologs from at least one parent. T1–T3: gain of AS in one or both homeologs. (b) AS events were observed in both diploids at a given gene, and the AS pattern changes post-polyploidization: gains or losses of AS in the homeologs from the different parents. T1: no change in homeolog AS patterns; T2: gain of novel AS isoform; T3–T5: loss of AS in one or both homeologs. Figure modified from Yoo et al. (2014a).

Moreover, the AS changes in *A. thaliana* and rice were reported to be affecting about 33% of genes (Filichkin et al. 2010; Zhang et al. 2010a).

Initially, AS patterns may be more stochastic and stability may be achieved over time, possibly resulting in AS patterns more like the parental patterns (Zhou et al. 2011). In view of that, a number of studies have been conducted on natural and synthetic polyploids to study AS changes (Saminathan et al. 2014; Zhou et al. 2011; Wang et al. 2018). For instance, Zhou et al. (2011) studied 82 AS events through sequencing and RT-PCR in *B. napus* polyploids and reported that various homeologous pairs exhibited

different AS patterns in natural allotetraploids. Resynthesized allotetraploids showed loss of a high fraction of AS events compared to the parents. On the contrary, an increase in AS levels has been reported in vegetative tissues of tetraploid watermelon compared with the corresponding diploid parent (Saminathan et al. 2014). A study conducted on natural allotetraploid cotton revealed that more than half of the homeologs in each sub-genome produced divergent transcriptional isoforms (Wang et al. 2018). All these findings demonstrate that polyploidy induced AS changes are frequent and may present many remarkable functional consequences in polyploids.

Alternative splicing in conjunction with transcriptional regulation controls tissue-specific expression of gene isoforms (Kornblihtt et al. 2004). Many observations have proposed that tissue-specific cues regulate several AS events (Yoshimura et al. 2002). For example, Shen et al. (2014) demonstrated that one third of the AS events displayed differential tissue expression in paleopolyploid soybean and different tissues exhibit dramatic variations in the frequency of AS types and number of AS events. Developing tissues like young seeds and the shoot meristem showed higher AS frequency as compared to other tissues (Shen et al. 2014). It might be possible that such tissue-specific AS events regulate in a synchronized manner in specific interaction pathways as groups of genes that are transcriptionally co-regulated (Blencowe 2006). Additionally, sub-functionalization of AS could also occur due to partitioning of AS forms between duplicated genes. For instance, in resynthesized *Brassica* allopolyploids, AS might have played an important role in the regulation of *PASTICCINO2 (PAS2)* expression level through division of parental AS patterns between homeologs. Homeologs derived from *B. rapa* and *B. oleracea* displayed retention of only intron 2 and intron 8, respectively (Zhou et al. 2011). This partitioning of AS forms could result in retention of both genes, if different functions are performed by each splice form.

Another possibility could be the degradation of intron-containing transcript isoforms through non-sense mediated RNA decay (Reddy 2007), which would potentially lower the total level of gene expression.

The spatiotemporal specificity of AS is governed by multiple factors, collectively termed the "splicing code", which work in a concerted manner (Wang and Burge 2008; Reddy et al. 2012). These include trans-regulatory elements (Busch and Hertel 2012) and cis-regulatory elements like intronic splicing silencers and enhancers (Yeo et al. 2007) and exonic splicing silencers and enhancers (Fu and Ares Jr 2014). Therefore, changes in AS patterns can occur when a combination of these diverged AS factors derived from both progenitor species leads to different interactions with target splice sites in genes (Riddle and Birchler 2003). Another possible mechanism can be the epigenetic changes in allopolyploids such as histone modifications and cytosine methylation changes, which might play a role in modulation of AS (Zhou et al. 2011). Moreover, genetic recombination (GR), transcriptional level and gene structure (intron length and GC content) may also affect AS, as reported in a *Brassica* hexaploid (Wang et al. 2019). AS genes comprise lower GC content, higher GR rates, more exons, a high level of expression and considerably longer introns (Shen et al. 2014). In short, allopolyploidization affects a selection of splice sites which results in gain and loss of AS events that subsequently lead to diverged AS patterns among allopolyploids and their parents. Such divergence in AS patterns may contribute to functional evolution of genes, which enables allopolyploids to become more adaptable to the environment.

1.6.4. Role of Post-Transcriptional/Translational Mechanisms in Shaping the Proteome

In cells, translation is one of the most energy consuming processes (Buttgereit

and Brand 1995; Lynch and Marinov 2015), therefore its proper regulation is necessary for controlling the level of gene expression and protein synthesis to the animals or plants actual needs. Considering additional genetic systems in the mitochondria and chloroplast and the complex translational apparatus of plants, translational regulation is particularly important in plants (Ferrando et al. 2017). Moreover, for the post-transcriptional mechanisms specifically, alternative splicing generates various functional patterns of gene expression and differences in abundance of transcripts that are not always meaningful biologically (Keene 2007; Joshi et al. 2011). Consequently, the correlation between mRNA abundance and protein levels is inconsistent, hence it is anticipated that proteins rather than transcripts modulate cellular biochemistry (Gygi et al. 1999). Thus, to increase working efficiency under various conditions, the interaction between various levels of gene regulation including co-transcriptional, post-transcriptional, and post-translational regulation is critical for plants (Reddy et al. 2013; Guerra et al. 2015; Skelly et al. 2016).

Alternative splicing (AS) may likely influence protein diversity and levels in plants. Different proteins can be translated from various alternatively spliced isoforms and could be found in a sample at different developmental stages and at different expression levels (reviewed in Reddy et al. 2013). In eukaryotic organisms, such proteins exhibit significantly diverse functional characteristics (Kelemen et al. 2013). That is, truncated proteins with lost functional domains (Keller et al. 2017), and proteins having altered subcellular locations (Shin et al. 2015) can be produced from spliced transcripts. Moreover, structurally and functionally diverse protein isoforms having gain or lost certain motifs might be translated from the mRNA variants (Zhang et al. 2019). Approximately 60-75% AS events are observed to occur in the protein-coding regions of mRNAs, and these events result in changes in protein stability, binding

properties, intracellular localization, enzymatic and signaling activities (Stamm et al. 2005; Campbell et al. 2006).

AS not only leads to increased protein diversity (Graveley 2001; Kazan 2003), but also influences the ratio of relative isoforms that further regulates protein levels (Wang et al. 2008). For instance, a study of up-frame shift mutants has proposed that AS potentially reduces the gene expression levels via nonsense mediated decay (NMD) that causes degradation of non-functional transcript isoforms having premature termination codons (PTCs) (Kalyna et al. 2011; Drechsel et al. 2013). It has been proposed that approx. one third of AS events could lead to the production of PTCs (Lewis et al. 2003; Wang and Brendel 2006), and thus are targeted for degradation by NMD (Kalyna et al. 2011; Drechsel et al. 2013; Ottens and Gehring 2016). As a result, phenotypic changes regarding physiological processes and developmental growth can be modulated by AS (Staiger and Brown 2013). Therefore, considering the role of AS in modulating protein levels and properties, it is regarded as an important mechanism that leads to proteome and transcriptome diversification, especially in mammals (Keren et al. 2010; Bush et al. 2017).

The role of AS in the development of functional protein diversity is less clear in plants as compared to humans (Kim et al. 2006). In polyploids, only a few studies have demonstrated expression at the protein level (Albertin et al. 2006; Albertin et al. 2007; Hu et al. 2011; Hu et al. 2013; Yao et al. 2011) and because of biological and technical limitations these studies have revealed data on homeolog-specific expression concerning only a few genes (Hu et al. 2011; Hu et al. 2013; Koh et al. 2012). These studies showed that polyploids may either display novel proteins not found in either parent or imbalanced contribution of parents to protein expression levels (reviewed in Soltis et al. 2014a). However, association between protein synthesis and transcript

abundance still needs to be revealed and in the absence of in-depth proteomic studies, the role of AS toward protein diversity is tenuous. A study conducted on maize and *Arabidopsis* depicted a negative association between translational efficiency and divergence of mRNA abundance, revealing that the paralog copy with lower translational efficiency has a higher abundance of mRNA (Wang and Chen 2019). Hence, regulation of duplicated genes at the translational level frequently contradicts, rather than follows, the divergence in abundance of mRNA, which somewhat compensates for their divergence in mRNA abundance (Wang and Chen 2019). Coate et al. (2014) reported that changes in translation varied soon after polyploidization to minimize expression differences between a *Glycine* polyploid and its parents. This, along with a study conducted by Wang and Chen (2019), unveiled the extensive impact of translational buffering on expression of duplicate genes in maize and *Arabidopsis*, suggesting that the degree to which function and expression of duplicated genes deviate is probably miscalculated when examined only at the transcriptional level. As functional similarity can be indicated by expression similarity (Blanc and Wolfe 2004; Wagner 2005), buffering of expression divergence at the post-transcriptional level might lessen the functional divergence among paralogs. Moreover, for duplicated genes, particularly those involved in the same pathway or in the same complex, it might be crucial to maintain the right dosage balance (Edger and Pires 2009; Birchler and Veitia 2012; Wang and Chen 2018; Gout et al. 2019). So, by buffering the expression divergence at the transcriptional level, the expression level of both paralog copies might be adjusted by translational regulation and so help them in maintaining appropriate gene dosage. Because protein amount and mRNA abundance are poorly associated (Gygi et al. 1999), it seems to be an interesting concept to reveal the evolutionary imbalance in allotetraploids. Therefore, proteomic approaches establish an alternative strategy to

examine if some metabolic networks or functional categories are under or over-represented, or modification of subcellular localization of the proteins is carried out in an allopolyploid.

Another significant mechanism causing variation at the protein level is post-translational modifications (PTMs). PTMs play crucial roles in the cell as they regulate protein subcellular location, turnover, and activity, or protein and protein/nucleic acids interactions and can be examined only at protein level (Uversky 2015, 2016). A few proteomic studies have described differential regulation of PTMs (Vincent et al. 2005), but with regards to allopolyploids limited data are available. In allopolyploids, proteomic data suggested that PTMs of the same gene product could be regulated differentially. For example, different amounts of protein isoforms (PTMs, but also homeo-alleles or paralogous genes) in synthetic *Brassica napus* may lead to novel metabolic equilibrium, contrary to its diploid parents (Albertin et al. 2007). Hence, protein abundance is the net product of translation, transcription, and other post-transcriptional/translational regulatory steps such as PTMs and mRNA degradation and therefore does not predict how these steps contribute individually. Thus, although regulation at the translational level possibly represents evolutionary and biologically important aspects of gene expression, numerous expression studies conducted on polyploids have only examined transcript abundance, leaving open the query of how regulatory steps, particularly PTMs and AS downstream of transcription, affect periodic patterns of gene expression.

1.7 Cytonuclear Interactions

The genetic control of plant's important metabolic pathways, such as photosynthesis and cellular respiration, is regulated by nuclear and cytoplasmic (i.e.,

plastid and mitochondrial) genomes. Because mitochondria and plastids originated via an endosymbiotic event, they still retain the remnants of their ancestral genomes from their respective alpha-proteobacterial and cyanobacterial progenitors, which resulted in a genomic division of labour (Gray and Archibald 2012a). It is noteworthy how multiple genomic compartments along with their complex arrangements remain preserved over several hundred millions of years of evolution. The genetics of cytoplasmic and nuclear genomes vary in almost every aspect including mutation rates, mechanism of expression and replication, copy number and mode of inheritance (reviewed in Sloan et al. 2018b). However, they still function in a concerted manner and profoundly affect the biology of the plant. Coevolution and integration among these genomic compartments is required for the execution of core eukaryotic functions (Rand et al. 2004). One of the important outcomes of cytonuclear integration is the presence and maintenance of enzyme complexes that are comprised of interacting subunits encoded by nuclear and cytoplasmic genomes. These 'chimeric' complexes play central roles in eukaryotic bioenergetics as they contain enzymes involved in photosynthesis, organelle ribosomes and cellular respiration (Rand et al. 2004). Their evolutionary history is reflected by this organization as numerous genes, ancestrally present in cytoplasmic genomes, have been replaced by gene transfer to the nucleus or substituted by existing nuclear genes (Sloan et al. 2018b). Besides these chimeric enzyme complexes, many nuclear encoded proteins interact either directly with the cytoplasmic genomes or their RNAs to mediate DNA replication, repair, transcription (Zhang et al. 2016; Gualberto and Newton 2017) and many post-transcriptional processes including intron splicing, tRNA aminoacylation, base modification, RNA editing and transcript end-processing (Germain et al. 2013; Salinas-Giegé et al. 2015). Hence, it is expected that in plant cells, variation in one genomic compartment can influence evolution in other genomes (Sloan 2015).

In addition to the aforementioned transcriptomic and genomic alterations that occur in polyploids, cytonuclear accommodation appears to be an important yet understudied aspect of allopolyploid evolution, considering the complex coordination between nuclear, mitochondrial and plastid genomes (Taylor 1989; Leon et al. 1998). During speciation of allopolyploids, two nuclear genomes from both progenitor species along with cytonuclear genomes from only one parent (usually the maternal) are maintained. Thus, two divergent regulatory hierarchies become combined and the stoichiometry between nuclear and organellar genes is altered, leading to physiological disruptions (reviewed in Gong et al. 2012). To achieve stability, selection is likely to favor compensatory mechanisms for maintaining coordinated expression between nuclear and cytoplasmic genes immediately after polyploidization (e.g., up-regulated cytoplasmic genome copy number, elevated levels of organelle biogenesis and down-regulation of expression of nuclear genes encoding organelle-targeted proteins) as well as over longer time periods (e.g., pseudogenization, sub-functionalization) (Sharbrough et al. 2017; Fernandes Gyorfy et al. 2021). Though gene balance theory (Birchler and Veitia 2007, 2010) and observations regarding cytoplasmic male sterility in interspecific hybrids (Schnable and Wise 1998) suggested that cytonuclear coevolution is a remarkable feature of evolution and responsible for stabilization of allopolyploid lineages, comparatively very little work has been done on the cytonuclear dimension of polyploid evolution so far.

Altered cytonuclear interactions, specifically plastid-nuclear incompatibilities, have been reported in allopolyploid species and interspecific hybrids (reviewed in Sharbrough et al. 2017). Several mechanisms may lead to these alterations, including the presence of non-synonymous substitutions that modify protein structure and regulatory networks (Greiner et al. 2011) and deletions in plastid intergenic sequences (Greiner et

al. 2008). Plastid-nuclear conflict is thought to be avoided by preferential transcription of nuclear genes inherited from the maternal parent over the paternally inherited copies. Thus far, it has been investigated in limited allopolyploid plant systems and mainly in genes encoding Rubisco (Ribulose 1, 5-bisphosphate carboxylase/oxygenase) subunits: small (SUU) and large (LSU) subunits (reviewed in Sharbrough et al. 2017). For instance, in *Gossypium hirsutum*, non-reciprocal homeologous recombination events were observed to occur between sub-genomes of natural allopolyploids, possibly leading to homogenization through maternal to paternal gene conversion between nuclear homeologs (Gong et al. 2012). Interestingly, preferential expression of maternal Rubisco homeologs was observed in both cultivated and wild cotton allopolyploids and also in resynthesized F_1 hybrids as compared to paternal homeologs (Gong et al. 2012). Likewise, between homeologous genomes of *Arachis hypogaea*, *Nicotiana tabacum* and *Arabidopsis suecica* allopolyploids, differential evolution of paternal and maternal Rubisco genes was reported (Gong et al. 2014). Accordingly, it has been proposed that interacting cytoplasmic genes might rapidly evolve due to their higher mutation rates or inefficient selection against deleterious mutations, which leads to subsequent changes in their nuclear counterparts in enzyme complexes (Osada and Akashi 2011; van der Sluis et al. 2015). This perspective suggests that cytonuclear co-evolution is primarily driven by cytoplasmic genomes, while nuclear genomes are viewed as primary responders. Even though there is mounting evidence regarding this model, many of its key components need to be tested critically.

Studies reporting contrasting results to maternal bias have also been published. The process of nuclear *rbcS* homogenization through gene conversion was not reported in resynthesized allotetraploid *Oryza sativa* or in young allopolyploid *Tragopogon miscellus*, although in these species Rubisco plastid-encoded parental sequences showed

non-synonymous substitutions (Sehrish et al. 2015; Wang et al. 2017b). Furthermore, Sehrish et al. (2015) revealed that cytonuclear coordination does not occur instantaneously upon polyploidy formation, as a low percentage of naturally occurring *T. miscellus* individuals exhibited maternal bias either in gene loss or expression. Additionally, an exceptional case occurs in *B. napus*, which inherited maternal diploid homeologs with no amino acid divergence from the paternal diploid homeolog (Gong et al. 2014). Based on this observation, it can be assumed that selection likelihood at the level of SSU/LSU interaction would be eliminated, and random interactions between homeologs regardless of their parental origin, can be expected. However, only intergenomic maternal to parental conversions were detected in *Brassica* (Gong et al. 2014). In support to this is the fact that various cytoplasmic factors are required to recognize SSU proteins, and these proteins need to be transported to the surface membrane of maternally derived plastids and subsequently, face the transmembrane channel for transport into the plastid. Therefore the gene conversion detected in *B. napus* reveals that selection may act at this level during some process of maternal trans-membrane transport (Bruce 2000; Lee et al. 2002). Besides this, the paternal homeolog with no maternal gene conversions showed expression bias in *Brassica* (Gong et al. 2014). Two possible situations can explain such diverse patterns of cytonuclear interactions in *B. napus:* 1) in plastids, weak selection of those maternal LSU that are identical to the paternal one; and/or 2) insufficient time for transcriptional selection to occur (Gong et al. 2014). Another study performed on *B. napus* involving different plastid enzyme complexes, reported no evidence of homogenization via maternal gene conversions and biased transcription in allopolyploids (Ferreira de Carvalho et al. 2019). These authors also described that even after ancient allopolyploidization, nuclear genes associated with plastid protein complexes, contrary to other plant protein complexes, likely have a

tendency of retaining genes as duplicate and triplicate over longer periods. Lastly, a recent study by Grover et al. (2022), conducted on large number of nuclear genes targeted to cytoplasmic genomes, did not show strong evidence of maternal bias to balance cytonuclear interactions in the examined allopolyploid systems. Such conflicting outcomes stated above suggest that systematic understanding of the mechanisms that uphold cytonuclear interactions in allopolyploids requires detailed research to include various enzyme complexes and pathways involved in functioning of plastids and mitochondria. More generally, polyploidy and cytonuclear interactions contribute towards the species innovation and plant diversity, thus, critical analysis of the phenomena that stabilize cytonuclear coordination in the face of whole genome duplications will eventually advise about plant evolution and diversification.

1.8 Dual-Targeted Proteins as Gene Expression Regulators

In plant cells, the energy transducing pathways connected to photosynthesis and respiration involve multimeric enzyme complexes that are comprised of both organelle and nuclear-encoded polypeptides and diverse cofactors whose synthesis and assembly must be well regulated and capable of responding to external and developmental stimuli (Pogson et al. 2008). Therefore, it is believed that precise communication and signaling pathways tend to exist between organelles (plastid, mitochondria) and the nucleus. Certainly, the nuclear genome encodes a large fraction of proteins necessary for organelle gene expression (OGE), thus allowing direct control of OGE via nucleus encoded factors (“anterograde control/signaling”). On the contrary, “retrograde signaling” refers to the concept that mitochondria and chloroplasts communicate messages to the nucleus through specific signaling molecules, thereby altering nuclear gene expression (NGE). Consequently, changes in physiological and developmental states of organelles can lead

to the modulation of NGE (reviewed in Kleine et al. 2009; Leister 2012; Pogson et al. 2008).

Dual-localization of proteins (in both the nucleus and organelles) makes them suitable candidates for their likely involvement in communication scenarios between the nucleus and organelles (Giegé and Duchêne 2012). Dual-targeted proteins are encoded by single genes and localized into two cellular compartments. As a result, both of the isoforms exhibit identical sequences in most of their length except for slight variation in their extremities because of the presence of targeting sequences (Giegé and Duchêne 2012). Thus far, dual-targeting has been documented in at least 15 diverse species of plants (Carrie and Whelan 2013). Surprisingly, in every species studied, the same proteins are found to be dual-targeted. For example, glutathione reductase has been examined to be dual-targeted in the flowering plants *Oryza sativa* (Xu et al. 2013), pea (Creissen et al. 1995), *Arabidopsis* (Chew et al. 2003) and the moss *Physcomitrella patens* (Xu et al. 2013). Among the known dually targeted proteins, only a few are recognized for their verified subcellular distribution in the nucleus and organelles (either mitochondria or chloroplast) (Krause et al. 2009). Amazingly, all of them appear to contribute to the maintenance of DNA and gene expression in the nucleus (Krause and Krupinska 2009). These proteins exhibit various targeting mechanisms. In general, these approaches are divided into two main categories (Krause and Krupinska 2009): the dual targeting of *de novo* proteins (Figure 1.5) and the relocalization of mature proteins through organelle–nucleus shuttle. In the latter case, proteins might either be attached to the organelle surface or are stored inside the organelles (Figure 1.6) (Krause and Krupinska 2009).

De novo synthesized proteins are dually targeted by different mechanisms (Figure 1.5). The most common mechanism is the use of multiple translation or transcription start sites to obtain multiple proteins from one gene, it can also include the

use of alternative exons through differential splicing (Krause and Krupinska 2009). In *Arabidopsis*, DNA ligase 1 demonstrates differential translation and transcription leading to dual-targeted nuclear and mitochondrial isoforms (Sunderland et al. 2006). On the contrary, the use of an alternative first exon through differential splicing is probably the reason for the nuclear and chloroplast isoforms of the plant RNA binding protein cp31 (Ohta et al. 1995). Another familiar mechanism for controlling the localization of those proteins with twin targeting motifs is post-translational modification of target proteins via for example, phosphorylation (reviewed in Karniely and Pines 2005; Silva-Filho 2003).

Figure 1.5: Mechanisms of dual-targeting of de novo synthesized proteins. Dual-targeting can be achieved by the use of two translation products or a single translation product. The use of two translation products can either be the result of alternative transcription initiation or differential splicing resulting in two separate mRNAs encoding different targeting signals. Alternative translation initiation can also be utilized to produce two different proteins with different targeting signals. Single translation products can be dual-targeted in two ways. The first uses two separate targeting signals on different ends of the protein and the second uses an ambiguous targeting signal which directs the protein to two organelles (Carrie and Whelan 2013).

In *Arabidopsis*, farnesylation (addition of a farnesyl group to a cysteine residue) was

shown to regulate the distribution to the nucleus and to plastids of an adenosine phosphate-isopentenyl-transferase (*AtIPT3*). The farnesylated GFP-tagged form of protein was observed to locate to the nucleus, while non-farnesylated protein was located inside the plastid (Galichet et al. 2008).

An alternative mechanism is the release of nuclear regulators, stored inside the organelles under certain conditions (Figure 1.6) (Krause and Krupinska 2009). It is proposed that nuclear gene expression can be affected significantly through nuclear regulators released from only a few of the many mitochondria and plastids (Krause and Krupinska 2009). Most likely, the evolution of organelles via the endosymbiotic event

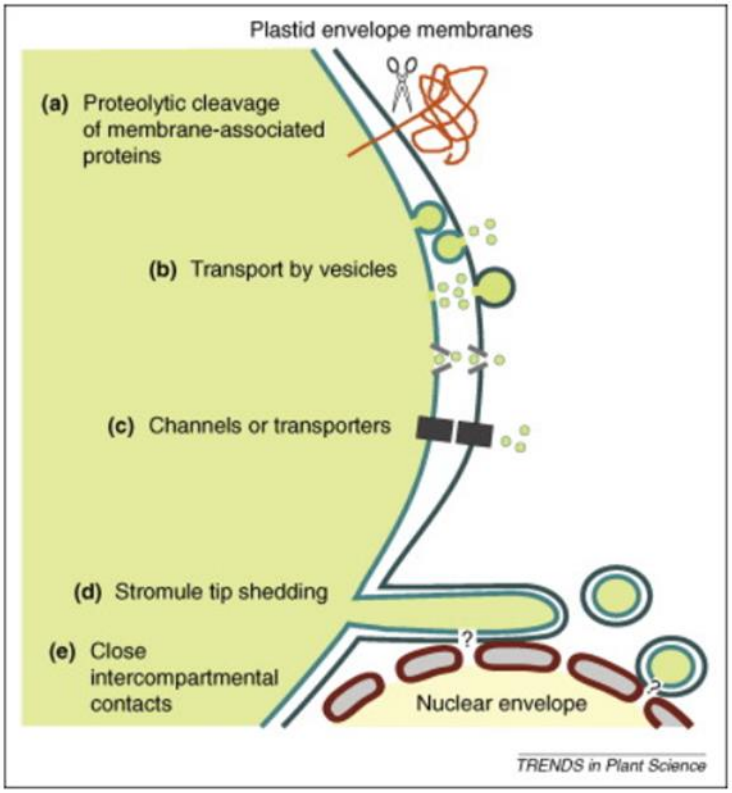


Figure 1.6: Schematic representation of different mechanisms of protein release from the organelles and shuttling to the nucleus: (a) release of protein from the organelle surface; (b) vesicle-mediated efflux of proteins from the organelle into the cytoplasm; (c) channels or transporters in the envelope membranes; (d) protein release via vesicles produced by tip shedding of stromules; and (e) close intercompartmental contacts. The question mark indicates that the mechanism is not known (Krause and Krupinska 2009).

from prokaryotic ancestors results in the preservation of transport systems for the release of stored proteins. In plants, components of the bacterial transport systems: the general secretory (Sec) pathway and the twin-arginine translocation (Tat) pathway have been observed only in the proteome of thylakoid membrane while not in mitochondrial or plastid envelopes (Krause and Krupinska 2009; Tjalsma et al. 2000). Nonetheless, retrograde translocation has been documented. For instance, mitochondrial fumarase protein might release from the organelle, shortly after the cleavage of the N-terminal target sequence in the mitochondrial matrix and enter the nucleus (Singh and Gupta 2006). In plants, numerous proteins from the *Arabidopsis* ABC protein directory are designated as candidates for plastid-envelope-based transporters based on their sequence analysis (Sánchez-Fernández et al. 2001). Many of these proteins have not been characterized yet, therefore, whether few accomplish an unknown function in retrograde protein translocation has not been established, compared to observations in animal and yeast mitochondria. Protein trafficking between compartments might also be done through transport vesicles similar to that seen in bacteria (Soltys and Gupta 1996; Singh et al. 1997; Radhamony and Theg 2006) or intercompartmental contacts (Soltys and Gupta 1999; Andersson et al. 2007; Seguí-Simarro et al. 2008).

Retrograde regulatory pathways from the chloroplast to the nucleus have been well studied in plants. These involve the control of nuclear gene expression via epistasy of synthesis and tetrapyrrole biosynthetic pathway (Pogson et al. 2008). It has been suggested that mitochondrial retrograde signaling to the nucleus includes reactive oxygen species (ROS) that are also characterized as mitochondrial oxidative stress markers (Rhoads and Subbaiah 2007). These retrograde regulatory pathways suggest that gene expression seems to be regulated by plant nucleo-organelle proteins, as done by their other eukaryotic counterparts. Therefore, these proteins serve as good candidates for

maintaining coordination of gene expression between the nucleus and organelles (Giegé et al. 2005; Giraud et al. 2010), which is necessary for organelle biogenesis and normal plant development. Nonetheless, no study has been published regarding the role of dual-targeted proteins in regulating gene expression in allopolyploids. Even though dual-targeted proteins are fewer in number compared to the total organelle proteomes, exciting questions related to inter-organelle communication are raised by the mechanism of dual-targeting. Thus, an understanding of the process of dual-targeting in the context of allopolyploidization may provide useful insights into the targeting of location-specific proteins to mitochondria or chloroplasts and their role as gene expression regulators.

1.9 *Tragopogon* as a Study System

The genus *Tragopogon* (Asteraceae) offers a unique system to study recurrent and recent allopolyploidy in natural populations. This system represents a typical example of instant speciation via polyploidization. In the early 1900s, the introduction of three diploid species: *Tragopogon pratensis, T. porrifolius* and *T. dubius* from Europe to western North America, subsequently led to the formation of two allotetraploid species (*T. miscellus* and *T. mirus*) (Ownbey 1950). DNA, flavonoid and isozyme studies were conducted to confirm the ancestries of both allotetraploid species (Ownbey and McCollum 1953, 1954; Brehm and Ownbey 1965; Roose and Gottlieb 1976; Soltis and Soltis 1989; Soltis et al. 1995b). *T. porrifolius* and *T. dubius* are the progenitors of *T. mirus*, while *T. dubius* and *T. pratensis* are the diploid progenitors of *T. miscellus* (Figure 1.7). Molecular data indicate that in just the past 90-100 years, *T. mirus* and *T. miscellus* have formed frequently, possibly 13 and 21 times, respectively (reviewed in Soltis et al. 2004). *T. miscellus* is of particular importance as it has reciprocal forms in nature that can easily be distinguished morphologically: *T. miscellus* exhibits short-

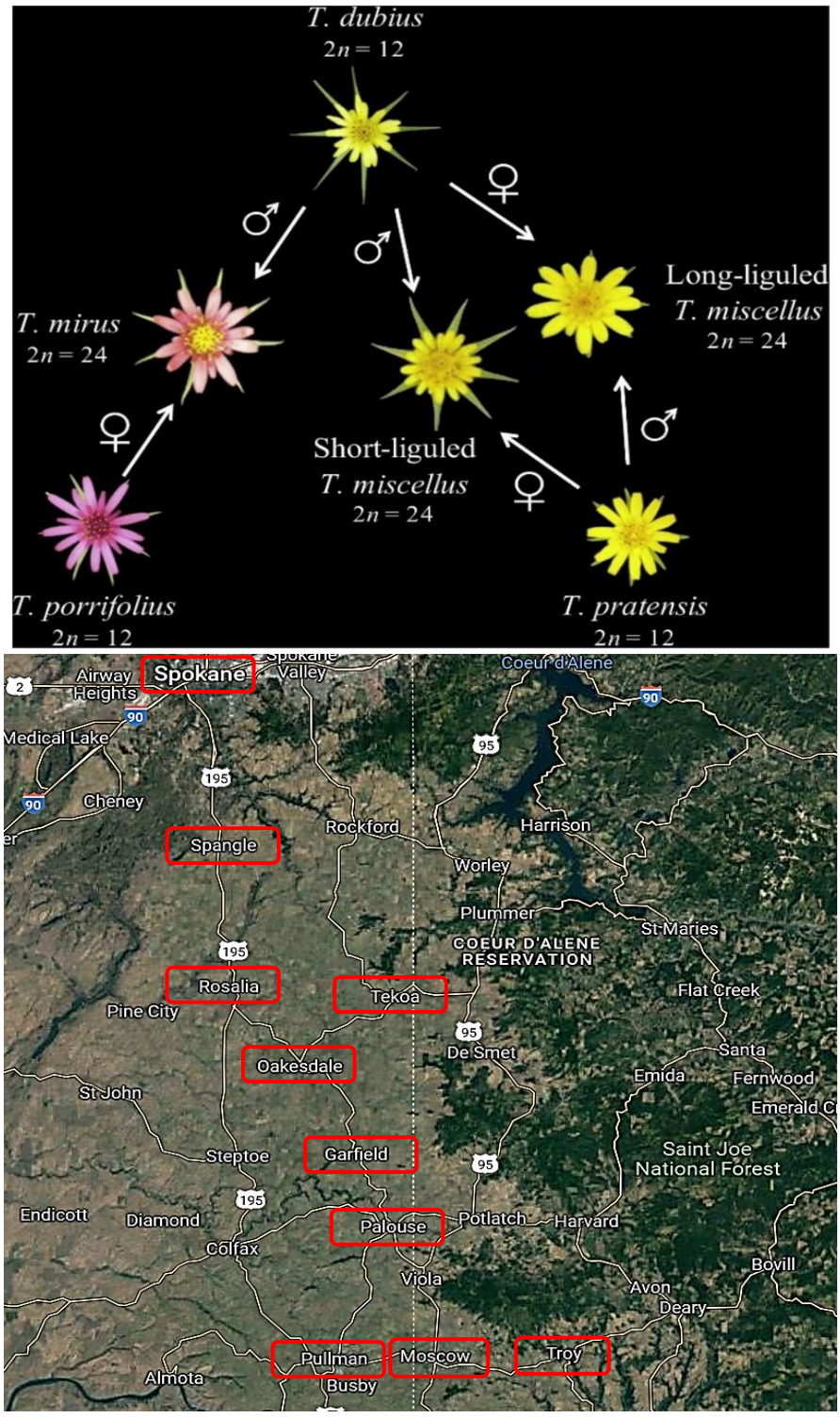


Figure 1.7: Allotetraploids of *Tragopogon* species and their ancestry on the top (Shan et al. 2020c). *T. porrifolius* and *T. dubius* are the parents of allotetraploid *T. mirus*. *T. pratensis* and *T. dubius* are the parents of reciprocally formed allotetraploids of *T. miscellus*. Long- and short-liguled forms exist when *T. dubius* and *T. pratensis* are the maternal parents, respectively. The map on the bottom shows USA towns (encircled red) where natural populations of parental species and reciprocal forms of *T. miscellus* allopolyploids are.

ligulate ray flowers when the maternal parent is *T. pratensis*, while *T. miscellus* shows

long-ligulate ray flowers when the maternal parent is *T. dubius* (Ownbey 1950; Wendel 2015). All existing populations (~40) of *T. miscellus* are short-liguled except for one that is long-liguled (Pullman, WA) (Novak et al. 1991).

Repeated formation of *Tragopogon* allopolyploids, including reciprocal forms (Soltis and Soltis 1995; Soltis et al. 2004; Soltis and Soltis 1991; Symonds et al. 2010) provides a way to determine if such independently formed allopolyploid populations develop similar cytonuclear coordination. Studies conducted previously have identified transcriptomic and genomic modifications in *Tragopogon* allopolyploids shortly after their formation, including homeolog loss and silencing, differential expression of homeologous loci, differential proteomes (Buggs et al. 2010b; Buggs et al. 2012a; Buggs et al. 2009b; Buggs et al. 2012b; Buggs et al. 2011a; Koh et al. 2012; Koh et al. 2010; Tate et al. 2009b; Tate et al. 2006a; Shan et al. 2020c) and extensive chromosomal variations, for instance, intergenomic translocations and aneuploidy (Chester et al. 2012b; Lim et al. 2008; Spoelhof et al. 2017). Moreover, genomic modifications at early stages of polyploid formation were also analyzed through the establishment of synthetic polyploids of *Tragopogon* (Tate et al. 2009c). Because recently formed natural polyploids provide a unique opportunity to investigate the immediate consequences of polyploidization, in this project the impact of the dynamic nature of polyploidy on cytonuclear interactions at molecular level was investigated using *Tragopogon miscellus* allopolyploids.

1.10 Research Project

The goal of this PhD project was to investigate cytonuclear interactions and coordination following whole genome duplications in *Tragopogon miscellus* allopolyploids. The focus was analysis of transcriptomic responses to cytonuclear

coordination in young, recurrent and reciprocally formed allopolyploids as compared to their parents. Below is a brief overview of the aims that were achieved during this project.

Chapter 2

Identification and Characterization of Cytonuclear Genes in *Tragopogon dubius*

This chapter involves the identification and characterization of nuclear-encoded organelle-targeted genes extracted from a draft *T. dubius* genome, as well as their counterpart organelle-encoded (mitochondrial and plastid) genes from previous high-throughput sequencing data. These genes are implicated in multi-subunit cytonuclear enzyme complexes (OXPHOS complexes, photosynthetic complexes, etc.) as well as dual-targeted genes. A significant portion of this thesis includes the investigation of cytonuclear interactions through gene expression analysis, hence, this chapter involves the first basic step in this regard. The genes identified in this chapter were used for investigation of cytonuclear interactions in the next chapter.

Chapter 3

Evolution of Duplicated Nuclear Genes (Homeologs) Implicated in Cytonuclear Interactions in *T. miscellus* Allopolyploids

To analyse transcriptomic responses to cytonuclear interactions, in this chapter, gene expression patterns of nuclear homeologs and their organellar counterparts involved in both mitochondrial oxidative phosphorylation and plastid complexes (identified in chapter 2), were investigated. In view of dual-targeting, three proteins: mTERF6, PNM1 and Whirly1 were examined for expression levels as well. RNA CaptureSeq (Targeted RNAseq) was used to evaluate the selected transcripts from RNA of reciprocally formed

natural *T. miscellus* populations and progenitor (*T. pratensis and T. dubius*) populations to determine the inherited expression differences between homeologs as well as novel expression patterns. CaptureSeq is a highly sensitive tool that involves hybridization of complementary oligonucleotides to enriched cDNA libraries for sequencing (Martin et al. 2016; Mercer et al. 2014). This method was preferred over traditional whole transcriptome RNA sequencing (RNAseq) because it focuses on evaluation of selected transcripts with increased sensitivity, allowing characterization of low or transiently expressed transcripts that may otherwise be missed using RNAseq. In addition, it can be done at reduced cost or scale, and thus permitted us to analyse 76 plant samples in this study.

Chapter 4

Investigation of Morphological Trait Variation in the Reciprocal *T. miscellus* Allopolyploids

In this chapter, morphological consequences of polyploidy were investigated in reciprocal forms of *T. miscellus* allopolyploids. This is a preliminary study, focussing on the variation in germination rates and leaf morphological traits (leaf number and length) in the same plants including both diploid parent species and *T. miscellus* allopolyploids that were used for gene expression analysis (Chapter 3).

Chapter 2

Identification and Characterization of Cytonuclear Genes in *Tragopogon dubius*

2.1. Abstract

The proper functioning of a plant's important bioenergetic processes requires tightly regulated and coordinated interactions among nuclear and cytoplasmic genomes. Both cytonuclear enzyme complexes and dual-targeted genes involve interacting nuclear and organellar genomic counterparts and are the key elements of cytonuclear interactions, providing a way to study cytonuclear co-evolution in a plant cell. In this study, putative nuclear and organelle genes involved in cytonuclear enzyme complexes, as well as dual-targeted genes, were identified from a draft *Tragopogon dubius* genome using the well-resourced angiosperm species, *A. thaliana* (Brassicaceae) and *L. sativa* (Asteraceae), as reference species. The purpose of doing so is ultimately to study the co-evolution and coordination of these complexes in allopolyploids of *Tragopogon*. We identified genes involved in six mitochondrial OXPHOS complexes, including an alternative OXPHOS pathway, nine plastid complexes and three dual-targeted genes. The subunit composition of the cytonuclear enzyme complexes in *T. dubius* varied compared to *A. thaliana* but was similar to *L. sativa*, which are in the same plant family, Asteraceae. The OXPHOS complex II in *T. dubius* species could not be identified in the nuclear genome and therefore could be cytonuclear as it is in *L. sativa*, where the *sdh3* and *sdh4* genes of complex II are mitochondrial-encoded. Because this putative list of cytonuclear genes was identified in a draft *T. dubius* genome, its validation from a complete annotated genome is needed, but was useful for characterizing the cytonuclear complexes and will

be used to design probes to further study the expression of the cytonuclear complexes in allopolyploids.

2.2 Introduction

The vast majority of genes in eukaryotic organisms, including plants, reside on the chromosomes in the nucleus. However, plants also have cytoplasmic genomes located in organelles, the mitochondrion and the chloroplast. According to the endosymbiotic theory, both mitochondria and chloroplasts are derived from the α-proteobacterial and cyanobacterial endosymbiotic ancestors, respectively (Gray and Archibald 2012b). During evolution, a large number of organellar genes have been transferred from ancestral organellar genomes to the nuclear genome, where they encode organelle-targeted proteins (Timmis et al. 2004; Sloan et al. 2018a; Weeden 1981). More than 90% of the organellar proteins are imported from the cytoplasm, suggesting the greater dependency of organellar genomes on the nucleus (Timmis et al. 2004). In animals, the endosymbiotic gene transfer to the nuclear genome no longer occurs, resulting in a stable mitochondrial genome. However, in flowering plants, gene transfer remains an active and recurring process (Timmis et al. 2004; Johnston and Williams 2016). For example, mitochondrial gene *rps10* and chloroplast gene *infA* have showed strong evidence of multiple and independent gene transfers to the nucleus (Adams and Palmer 2003; Adams et al. 2000; Millen et al. 2001). Therefore, in flowering plants, the tripartite distribution of genomes plays a significant role in their evolution and diversification (Roux et al. 2016; Sharbrough et al. 2017). For example, organelle-targeted proteins interact with the organelle-encoded proteins and form multi-subunit enzyme complexes. These cytonuclear enzyme complexes regulate various important processes including aerobic respiration and photosynthesis in both mitochondria and the chloroplast, respectively (Rand et al. 2004). The cytonuclear enzyme complexes are also involved in ribosome biogenesis in both organelles (Rand et al. 2004; Greiner and Bock 2013). In addition, nuclear-encoded organelle-targeted proteins also directly interact with the organelle

genomes to assist in DNA recombination, replication and repair, as well as transcriptional processes (Zhang et al. 2016; Havird et al. 2017). Thus, protein-protein and protein-DNA interactions among these genomes require dynamic and coordinated regulation for appropriate functionality of plants.

Cytonuclear coordination and co-evolution is influenced by the evolutionary differences between nuclear and organellar genomes, such as differences in mutation rates, mechanisms of replication and mode of inheritance (Sloan et al. 2018a). Notably plant mitochondrial and chloroplast genomes have different genome sizes and substitution rates. The mitochondrial genome generally has lower substitution rates than both the chloroplast and nuclear genomes (Wolfe et al. 1987). In the case of genome size, the chloroplast genome size (120-160 kb) is more conserved across most flowering plants (Wicke et al. 2011; Xiao-Ming et al. 2017) compared to the mitochondrial genome size (66 kb-11.3 Mb) (Sloan et al. 2012; Skippington et al. 2015). It is likely that mutation or evolutionary change in organelle genomes trigger compensatory changes in the corresponding nuclear genes, thus impacting the associated crucial mechanisms (Osada and Akashi 2012b; Burton et al. 2013; Rockenbach et al. 2016). For instance, the accelerated evolution of the chloroplast genome in *Silene* species resulted in the rapid evolution of nuclear-encoded subunits of caseinolytic protease and Acetyl Co-A carboxylase complexes, which physically interact with organelle-encoded subunits (Rockenbach et al. 2016). Therefore, cytonuclear co-evolution is an important phenomenon occurring in the life history of eukaryotes.

To understand cytonuclear coordination and molecular co-evolution in plants, both plastid and mitochondrial multi-subunit enzyme complexes, as well as dual-targeted genes, need to be identified and characterized. Dual-targeted genes are those genes whose protein products are targeted to more than one organelle in the cell (discussed in detail in

chapter 1, section 1.6). Cytonuclear enzyme complexes have been examined in many organisms and lineage-specific differences in the subunit composition have been noticed (Adams et al. 2002; Mower and Bonen 2009; Shrestha et al. 2020). For example, the number of mitochondrial OXPHOS complex I subunits varies among different species including plants (*Arabidopsis* 48 subunits), animals (cows 45) and fungi (*Yarrowia lipolytica* 40) (Carroll et al. 2006; Morgner et al. 2008; Forsythe et al. 2019). Within plant species, the variation in the accessory subunits of OXPHOS complex I also has been observed, leading to the differential subunit composition (Meyer 2012). Moreover, the occurrence of subunit isoforms (gene duplicates) as detected in the *Arabidopsis* genome could also lead to variable numbers of subunits across different species (Klodmann et al. 2010). In addition to these, in flowering plants, the active event of endosymbiotic organelle gene transfer to the nucleus can also cause alteration in the cytonuclear interactions across plant species (Adams and Palmer 2003; Timmis et al. 2004). For example, the OXPHOS complex II can be either cytonuclear or entirely nuclear-encoded in plant species due to the transfer of *sdh3* and *sdh4* genes (Timmis et al. 2004).

In the current study, the subunit composition of various cytonuclear enzyme complexes in *Tragopogon dubius* (Asteraceae), a diploid species that is a parent of the polyploid species *T. miscellus* and *T. mirus*, were determined. In this regard, nuclear- and organelle-encoded genes that are part of multi-subunit enzyme complexes were identified in a *T. dubius* draft genome. Moreover, the dual-targeted genes (targeting to chloroplast/mitochondria and the nucleus) were identified as well. In addition to the cytonuclear genes, the entirely nuclear-encoded control genes that are not involved in cytonuclear interactions, including housekeeping genes, retrograde-signalling partners of dual-targeted genes, as well as genes implicated in the OXPHOS alternative pathway, were also retrieved from the *T. dubius* genome.

2.3. Methods

2.3.1. Genomic Resources Used for Protein Identification

The CyMIRA (The Cytonuclear Molecular Interactions Reference for *Arabidopsis*) database was used to retrieve the list of nuclear genes involved in the candidate cytonuclear enzyme complexes as well as dual-targeted genes in *Arabidopsis thaliana* (Supplementary Table 2.1) (Forsythe et al. 2019). The list of control genes that are entirely nuclear-encoded and are not involved in cytonuclear interactions in *A. thaliana,* such as genes implicated in OXPHOS complex II, OXPHOS alternative pathways, housekeeping genes and retrograde-signalling partners of dual-targeted genes, was also retrieved (Supplementary Table 2.2) (Millar et al. 2004; Expósito-Rodríguez et al. 2008; Klodmann et al. 2011; Hammani et al. 2011; Senkler et al. 2017; Ren et al. 2017; Jin et al. 2019). To identify these nuclear genes implicated in cytonuclear interactions as well as control genes in *T. dubius* species, all nuclear-encoded proteins in the *A. thaliana* Araport11 genome annotation were used as a primary reference. Because extensive work on cytonuclear biology in this model angiosperm has been conducted, its genomic annotation can be extended to genomic datasets from non-model plant species to obtain their functional data. For candidate species, we used nucleotide and protein sequences available from a draft *T. dubius* genome (Soltis lab). Because the *T. dubius* genome is a draft (version 1) and some proteins were missing from its annotation, the nuclear-encoded proteins of the model Asteraceae, *Lactuca sativa* cv Salinas (Genome ID 28333, version 8), were also used as a secondary reference to identify these additional proteins.

2.3.2. Ortholog Identification of Organelle-targeted and Organelle Non-targeted Nuclear Proteins

The detailed overview of protein orthologs' identification is illustrated in Figure 2.1. Initially, the Proteinortho (version 5.16) tool was used to identify putative orthologs for all *A. thaliana* cytonuclear proteins (targeted to organelles) and selected control proteins (not targeted to organelles) in *T. dubius* (Figure 2.1). For this purpose, the

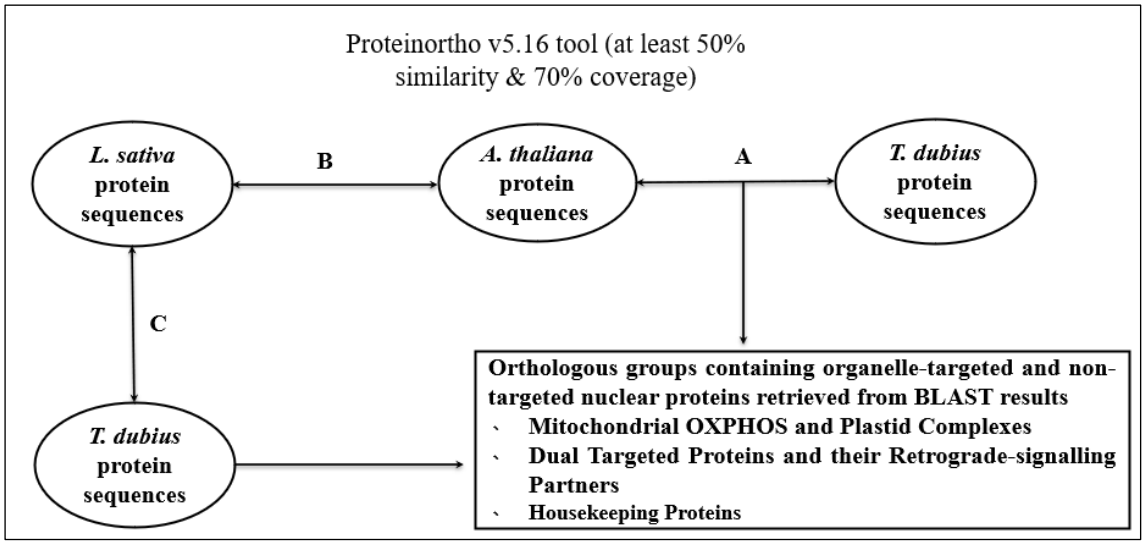


Figure 2.1 : Ortholog identification of organelle-targeted and organelle non-targeted nuclear proteins in the *Tragopogon dubius* genome. The first step involved an alignment of canonical isoforms of all *T. dubius* and *Arabidopsis thaliana* proteins (A), followed by an alignment between *Lactuca sativa* and *A. thaliana* proteins in the second step (B) to find missing *A. thaliana* proteins from first step. Lastly, protein orthologs corresponding to *Lactuca* proteins found in the second step were identified in the *T. dubius* genome (C).

canonical isoforms for all proteins from the *A. thaliana* genome and the *T. dubius* draft genome were used as input. Proteins were classified into orthologous groups with each member having at least 50% identity and 70% overlap. All identified orthologous groups containing one of the *A. thaliana* organelle-targeted nuclear proteins, including mitochondrial OXPHOS complexes, plastid complexes, dual-targeted proteins, and control proteins, including the OXPHOS alternative pathway, retrograde-signalling and housekeeping proteins, were classified into 1:1, 1:n, n:1 or n:n classes (Figure 2.2).

Category 1:1 is a protein orthologous group that includes only single-copy protein orthologs. Categories 1:n and n:1 included putative protein groups where we found protein paralogs/duplicates in *T. dubius* for an *Arabidopsis* protein or vice versa, respectively.

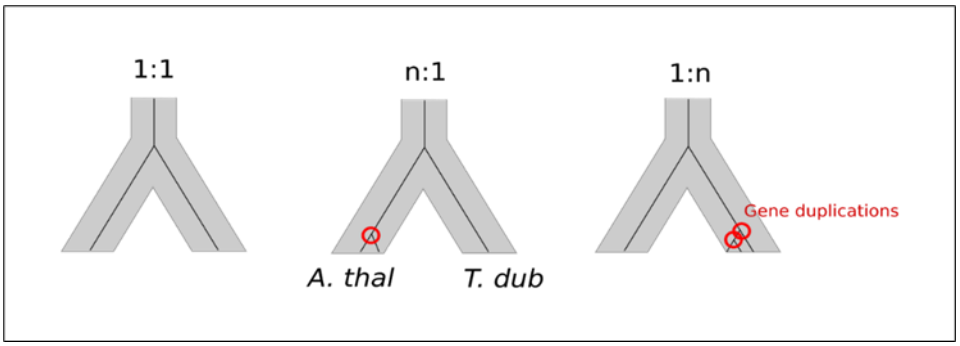


Figure 2.2: Representation of orthologous groups identified in the *Tragopogon dubius* genome. The grey area represents the organismal relationship between *Arabidopsis thaliana* and *T. dubius*. The black lines represent relationships among genes from each species. Category 1:1 represents orthologous protein groups that are single copy in both lineages. Categories n:1 or 1:n include gene duplications in *Arabidopsis* or *Tragopogon*, respectively.

Secondly, the orthologous groups for those *A. thaliana* proteins for which no ortholog was found in the *T. dubius* proteome were also identified using the canonical isoforms of the *Lactuca sativa* proteins via Proteinortho tool with the same BLAST and clustering parameters as described above (Figure 2.1).

Lastly, to proceed with identification of putative protein orthologs in *T. dubius*, the Proteinortho tool with the same BLAST and clustering parameters was again used: at least 50% identity and 70% overlap. All proteins from *L. sativa* and the *T. dubius* genome were utilized as input for this analysis (Figure 2.1). From the BLAST output, the selected *L. sativa* proteins retrieved in the previous step (BLAST between *L. sativa* and *A. thaliana* proteins) were used to identify their respective putative orthologs in *T. dubius*.

2.3.3. Screening for Single-copy Protein Orthologs

For putative proteins classified as 1:n, n:1, or n:n, the most likely single-copy orthologs were screened by comparing the length and percent-identity of all protein alignments. However, we could not differentiate between these putative proteins as their BLAST results showed that they are of almost similar length and have the same identity. Given a large number of proteins identified as putative homologs to *A. thaliana* light-harvesting proteins, the relationship among these proteins was examined more closely. The gene family that includes all *Arabidopsis* light-harvesting proteins was accessed from the Ensembl Plants database (https://plants.ensembl.org/index.htm) and the amino acid sequences of all proteins from this gene family in Brassicaceae and Asteraceae were retrieved. These sequences were aligned together with *T. dubius* proteins identified as putative homologs to the *Arabidopsis* light-harvesting proteins using MAFFT (version 7.310) (Katoh and Standley 2016). A phylogenetic tree was constructed using a maximum likelihood approach and the "CAT" model of amino acid substitution, as implemented in RAxML (version 8.2.11) (Stamatakis 2014) (Figure 2.3). This tree showed that light-harvesting complex proteins in photosystem II displayed higher homology with each other within the same species and form clades, suggesting these gene families may have undergone multiple rounds of gene duplications in each lineage. Considering this complexity, we concluded that these putative proteins are paralogs in *T. dubius*.

2.3.4. Retrieving Nucleotide Sequences for the Identified Protein Orthologs and Paralogs

Gene models were created for each *T. dubius, L. sativa* and *A. thaliana* organelle-targeted protein identified. Protein sequences were aligned to the respective genomes, including *T. dubius* draft genome, as well as *L. sativa* and *A. thaliana* genomes, using

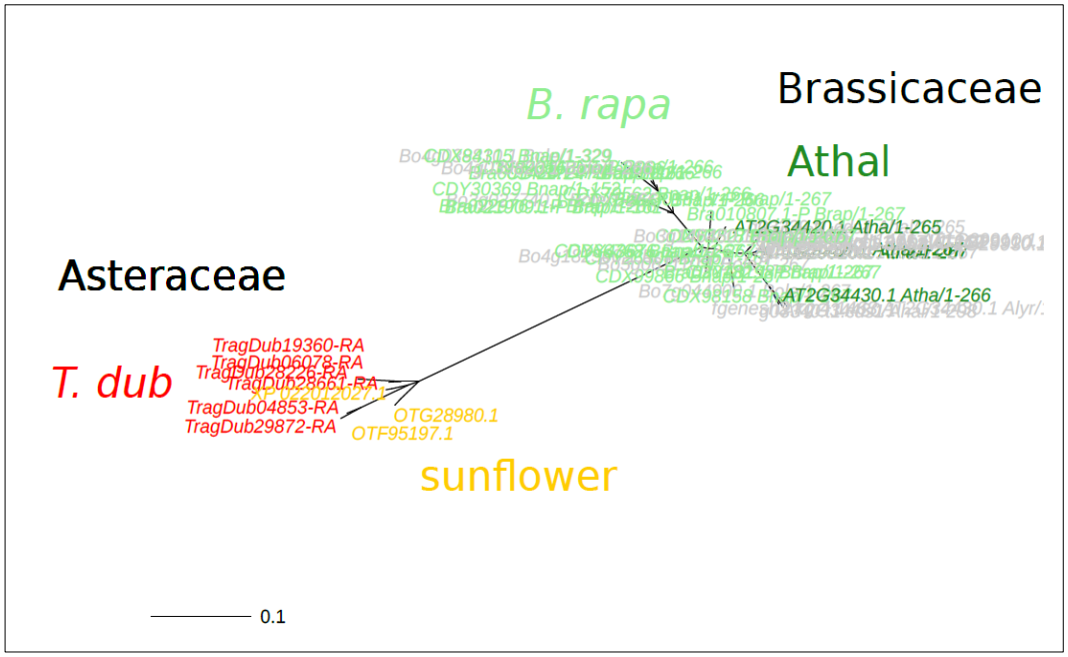


Figure 2.3: Light harvesting proteins form a complex gene family with many species-specific paralogs. A maximum likelihood tree was estimated from protein sequences. Tips are labeled with respect to species, *T.dub* = *Tragopogon dubius* (red), sunflower = *Helianthus annuus* (yellow), *B. rapa* = *Brassica rapa* (light green), *Athal* = *Arabidopsis thaliana* (dark green). Grey labels are from a variety of other Brassicaceae taxa.

exonerate (version 2.4) (Slater and Birney 2005) and the 'protein2genome' alignment model. Nucleotide sequences were extracted for each gene using the 'getfasta' utility from bedtools (version 2.30.0) (Quinlan 2014).

2.3.5. Identification of Organellar (Plastid and Mitochondrial) Genes

The genes for organelle-encoded counterparts of cytonuclear enzyme complexes were also identified. For plastid-encoded genes, an available annotated *T. dubius* plastid genome (J. Tate, unpublished data) was used to find sequences of genes involved in selected plastid enzyme complexes. In the case of mitochondrial-encoded genes, an annotated *Lactuca sativa* mitochondrial genome (NC_042756) was used to obtain sequences for genes implicated in mitochondrial OXPHOS complexes because we currently lack an annotated mitochondrial genome for *Tragopogon* species.

The identification of candidate organelle genes was done by importing the annotated *T. dubius* plastid and *L. sativa* mitochondrial genomes into Geneious (version 9.1.8) (Biomatters, Ltd.) and simply retrieving FASTA sequences for the candidate genes.

2.4. Results

2.4.1. Identification of Orthologs of Organelle-Targeted Proteins

In total 169 organelle-targeted putative orthologous proteins were identified in *T. dubius*, and 21 proteins were selected from *L. sativa* (Table 2.1). The *L. sativa* proteins are compared against *Arabidopsis* proteins for which no orthologous proteins were found in the *T. dubius* draft genome.

The additional details for the loci of organelle-targeted putative single-copy orthologs as well as paralogs are given in Tables 2.2-2.4. Few *T. dubius* proteins (highlighted in red) were identified using the selected *L. sativa* proteins. However, there were certain *Lactuca* proteins that gave no BLAST hit in the *T. dubius* draft genome (Tables 2.2-2.4). Likewise, the list of *Arabidopsis* proteins with missing orthologs in *T. dubius* and *L. sativa* genomes is presented in supplementary data (Supplementary Table 2.3).

Table 2.1: Number of putative protein orthologs identified in *Tragopogon dubius* and *Lactuca sativa.*

Complexes	**Complex names**	**Total proteins (*A. thaliana*)**	**Proteins retrieved from *T. dubius***	**Proteins retrieved from *L. sativa***	**Missing *A. thaliana* proteins**	**Total proteins (*T. dubius* & *L. sativa*)**
Mitochondrial OXPHOS Complexes	Complex I	48	38	4	1	42
	Complex II	13	5	2	4	7
	Complex III	14	8	0	3	8
	Complex IV	14	11	0	3	11
	Complex V	15	11	1	0	12
Plastid Complexes	ATP synthase	3	3	0	0	3
	Photosystem I	19	14	1	2	15
	Photosystem II	24	21	5	3	26
	Cytochrome b6f	2	1	1	0	2
	Rubisco	4	2	0	0	2
	Caseinolytic protease	15	12	2	1	14
	Acetyl-CoA carboxylase	4	3	0	0	3
	Ribosomal genes	42	32	5	3	37
	RNA polymerase PEP	6	5	0	1	5
Dual-targeted Genes	Nuclear-mito, Nuclear-plastid, or Nuclear-mito/plastid	3	3	0	0	3
Total		226	169	21	21	190

Table 2.2: List of identified putative proteins involved in OXPHOS complexes in *Tragopogon dubius* and *Lactuca sativa*. Red highlighted *T. dubius* protein was identified using the corresponding *L. sativa* proteins. The unique codes (e.g., At5g47890, Lsat_20989 and TragDub16379-RA) represent the associated genome loci of each protein in *Arabidopsis thaliana*, *L. sativa* and *T. dubius* genomes, respectively. Additional details of these proteins are given in Supplementary Table 2.1.

Enzyme Complex	***A. thaliana***	***L. sativa***	***T. dubius***
OXPHOS Complex I	At5g47890		TragDub16379-RA
OXPHOS Complex I	At3g12260		TragDub26927-RA
OXPHOS Complex I	At5g08060		TragDub06569-RA
OXPHOS Complex I	At2g20360		TragDub12281-RA
OXPHOS Complex I	At2g42210		TragDub23106-RA
OXPHOS Complex I	At3g03100		TragDub21105-RA
OXPHOS Complex I	At2g31490		TragDub29984-RA
OXPHOS Complex I	At2g02050		TragDub12762-RA
OXPHOS Complex I	At5g47570		TragDub07138-RA
OXPHOS Complex I	At4g34700		TragDub14659-RA
OXPHOS Complex I	At4g20150		TragDub10962-RA
OXPHOS Complex I	At5g37510		TragDub26636-RA
OXPHOS Complex I	At5g67590		TragDub10325-RA
OXPHOS Complex I	At3g03070		TragDub14746-RA
OXPHOS Complex I	At5g11770		TragDub04484-RA
OXPHOS Complex I	At5g08530		TragDub17396-RA
OXPHOS Complex I	At4g02580		TragDub03149-RA
OXPHOS Complex I	At1g67350		TragDub11786-RA
OXPHOS Complex I	At2g27730		TragDub05527-RA
OXPHOS Complex I	At1g67785		TragDub03625-RA
OXPHOS Complex I	At4g16450		TragDub08105-RA
OXPHOS Complex I	At4g00585		TragDub16685-RA
OXPHOS Complex I	At3g47930		TragDub24868-RA
OXPHOS Complex I	At1g68680		TragDub16873-RA
OXPHOS Complex I	At5g14105	Lsat_20989 and Lsat_24507[1]	TragDub06960-RA
OXPHOS Complex I	At3g08610		TragDub05616-RA and TragDub16818-RA[1]
OXPHOS Complex I	At5g52840		TragDub20259-RA and TragDub26190-RA[1]
OXPHOS Complex I	At2g33220 and At1g04630[1]		TragDub00026-RA and TragDub22160-RA[1]
OXPHOS Complex I	At2g02510 and At1g14450[1]		TragDub04873-RA
OXPHOS Complex I	At1g49140 and At3g18410[1]		TragDub01644-RA
OXPHOS Complex I	At2g42310 and At3g57785[1]		TragDub17966-RA
OXPHOS Complex I	At3g62790 and At2g47690[1]		TragDub05836-RA
OXPHOS Complex I	At1g79010 and At1g16700[1]		TragDub15230-RA
OXPHOS Complex I	At1g19580		TragDub22387-RA
OXPHOS Complex I	At1g47260		TragDub22387-RA
OXPHOS Complex I	At5g63510		TragDub01208-RA

Enzyme Complex	***A. thaliana***	***L. sativa***	***T. dubius***
OXPHOS Complex I	At3g48680		TragDub01208-RA
OXPHOS Complex I	At2g46540	Lsat_22998 and Lsat_8331[1]	
OXPHOS Complex I	At3g06310	Lsat_34290	
OXPHOS Complex I	At5g18800	Lsat_34290	
OXPHOS Complex I	At1g76200	Lsat_21648	
OXPHOS Complex II	At3g27380, At5g40650 and At5g65165[1]		TragDub21450-RA,TragDub28696-RA and TragDub19713-RA[1]
OXPHOS Complex II	At3g47833 and At5g62575[1]		TragDub02970-RA
OXPHOS Complex II	At5g66760 and At2g18450[1]		TragDub27665-RA
OXPHOS Complex II	At1g47420	Lsat_10503	
OXPHOS Complex II	At1g08480	Lsat_18951	
OXPHOS Complex III	At3g02090		TragDub27225-RA
OXPHOS Complex III	At1g51980		TragDub13028-RA
OXPHOS Complex III	At3g52730		TragDub06666-RA
OXPHOS Complex III	At5g40810 and At3g27240[1]		TragDub00409-RA
OXPHOS Complex III	At5g13430 and At5g13440[1]		TragDub00693-RA and TragDub14875-RA[1]
OXPHOS Complex III	At4g32470 and At5g25450[1]		TragDub01752-RA
OXPHOS Complex III	At3g10860 and At5g05370[1]		TragDub11010-RA
OXPHOS Complex IV	At1g80230		TragDub00413-RA
OXPHOS Complex IV	At4g21105		TragDub22246-RA
OXPHOS Complex IV	At1g72020		TragDub25997-RA
OXPHOS Complex IV	At1g22450		TragDub17478-RA
OXPHOS Complex IV	At2g16460		TragDub29090-RA
OXPHOS Complex IV	At4g00860 and At1g01170[1]		TragDub11388-RA
OXPHOS Complex IV	At3g62400 and At5g61310[1]		TragDub15618-RA
OXPHOS Complex IV	At4g37830		TragDub17875-RA and TragDub23702-RA[1]
OXPHOS Complex IV	At5g27760		TragDub19150-RA and TragDub20137-RA[1]
OXPHOS Complex V	At1g51650		TragDub03880-RA
OXPHOS Complex V	At5g47030		TragDub11991-RA
OXPHOS Complex V	At5g13450		TragDub23464-RA
OXPHOS Complex V	At3g52300		TragDub13305-RA
OXPHOS Complex V	At2g21870		TragDub22478-RA
OXPHOS Complex V	At5g08670, At5g08680 and At5g08690[1]		TragDub08803-RA
OXPHOS Complex V	At4g30010		TragDub03894-RA and TragDub06877-RA[1]
OXPHOS Complex V	At4g29480, At2g19680 and At4g26210[1]		TragDub18597-RA
OXPHOS Complex V	At3g46430 and At5g59613[1]		TragDub00280-RA and TragDub22194-RA[1]
OXPHOS Complex V	At2g33040	Lsat_10787	

[1] *Paralogous proteins*

Table 2.3: List of identified putative proteins involved in plastid complexes in *Tragopogon dubius* and *Lactuca sativa.* Red highlighted *T. dubius* proteins were identified using the corresponding *L. sativa* proteins. The unique codes (e.g., At3g54890, Lsat_20510 and TragDub17351-RA) represent the associated genome loci of each protein in *Arabidopsis thaliana*, *L. sativa* and *T. dubius* genomes, respectively. Additional details of these proteins are given in Supplementary Table 2.1.

Enzyme Complex	***A. thaliana***	***L. sativa***	***T. dubius***
Photosystem I	At3g54890		TragDub17351-RA
Photosystem I	At3g61470		TragDub17388-RA
Photosystem I	At1g61520		TragDub29220-RA
Photosystem I	At3g47470		TragDub03406-RA
Photosystem I	At1g19150		TragDub18644-RA
Photosystem I	At2g20260		TragDub04930-RA
Photosystem I	At1g31330		TragDub16190-RA
Photosystem I	At1g55670		TragDub03866-RA
Photosystem I	At1g30380		TragDub02303-RA
Photosystem I	At4g12800		TragDub23652-RA
Photosystem I	At5g64040		TragDub25070-RA
Photosystem I	At2g46820		TragDub24947-RA
Photosystem I	At4g02770 and At1g03130[1]		TragDub00390-RA
Photosystem I	At3g16140 and At1g52230[1]		TragDub10108-RA
Photosystem I	At1g08380	Lsat_20510	
Photosystem II	At1g06680		TragDub09720-RA
Photosystem II	At1g79040		TragDub03856-RA
Photosystem II	At1g44575		TragDub05540-RA
Photosystem II	At2g30570		TragDub01922-RA
Photosystem II	At2g06520		TragDub28773-RA
Photosystem II	At1g67740		TragDub03620-RA
Photosystem II	At5g54270		TragDub17408-RA
Photosystem II	At4g10340		TragDub03786-RA
Photosystem II	At1g15820		TragDub24331-RA
Photosystem II	At5g66570 and At3g50820[1]		TragDub03157-RA,TragDub14927-RA and TragDub26172-RA[1]
Photosystem II	At4g21280 and At4g05180[1]		TragDub24003-RA
Photosystem II	At1g29920, At1g29910,At1g29930, At2g34430 and At2g34420[1]		TragDub04853-RA,TragDub06078-RA,TragDub19360-RA,TragDub28226-RA,TragDub28661-RA,TragDub29872-RA[1]
Photosystem II	At5g01530		TragDub17135-RA and TragDub25925-RA[1]
Photosystem II	At3g21055	Lsat_9599	
Photosystem II	At2g05100	Lsat_10456, Lsat_11580, Lsat_8208 and Lsat_9332[1]	

Enzyme Complex	*A. thaliana*	*L. sativa*	*T. dubius*
RNA polymerase PEP	At1g64860		TragDub25278-RA
RNA polymerase PEP	At3g53920		TragDub02042-RA
RNA polymerase PEP	At5g24120		TragDub00311-RA
RNA polymerase PEP	At2g36990		TragDub12384-RA
RNA polymerase PEP	At1g08540	Lsat_21617	TragDub24770-RA
ATP Synthase	At4g04640		TragDub04991-RA
ATP Synthase	At4g09650		TragDub19438-RA
ATP Synthase	At4g32260		TragDub14417-RA
Cytochrome b6f	At4g03280		TragDub24280-RA
Cytochrome b6f	At2g26500	Lsat_32495	
Caseinolytic Proteases	At1g66670		TragDub12055-RA
Caseinolytic Proteases	At1g02560		TragDub27574-RA
Caseinolytic Proteases	At1g12410		TragDub24707-RA
Caseinolytic Proteases	At1g09130		TragDub20134-RA
Caseinolytic Proteases	At4g17040		TragDub14867-RA
Caseinolytic Proteases	At5g51070		TragDub10810-RA
Caseinolytic Proteases	At1g68660		TragDub26227-RA
Caseinolytic Proteases	At4g25370 and At4g12060[1]		TragDub14823-RA and TragDub23280-RA[1]
Caseinolytic Proteases	At5g45390		TragDub15219-RA and TragDub29528-RA[1]
Caseinolytic Proteases	At5g50920 and At3g48870[1]		TragDub10743-RA
Caseinolytic Proteases	At1g11750	Lsat_20401	
Caseinolytic Proteases	At1g49970	Lsat_36990	
Acetyl-CoA Carboxylase	At2g38040	Lsat_35936	TragDub05027-RA
Acetyl-CoA Carboxylase	At5g35360		TragDub14126-RA
Acetyl-CoA Carboxylase	At5g16390 and At5g15530[1]		TragDub02230-RA
Rubisco	At1g67090, At5g38410, At5g38430 and At5g38420[1]		TragDub14370-RA and TragDub24397-RA[1]
Ribosomal Genes	At3g52150		TragDub09573-RA
Ribosomal Genes	At2g38140	Lsat_28510	TragDub22127-RA
Ribosomal Genes	At3g56910		TragDub02089-RA
Ribosomal Genes	At5g17870		TragDub21873-RA
Ribosomal Genes	At1g07320		TragDub21729-RA
Ribosomal Genes	At1g05190		TragDub25148-RA
Ribosomal Genes	At3g44890		TragDub02462-RA
Ribosomal Genes	At5g13510		TragDub09605-RA
Ribosomal Genes	At1g32990		TragDub19994-RA
Ribosomal Genes	At1g78630		TragDub26950-RA
Ribosomal Genes	At3g25920		TragDub03561-RA
Ribosomal Genes	At3g54210		TragDub04410-RA
Ribosomal Genes	At1g48350		TragDub16283-RA

Enzyme Complex	*A. thaliana*	*L. sativa*	*T. dubius*
Ribosomal Genes	At5g47190 and At4g17560[1]	Lsat_34702	TragDub21350-RA
Ribosomal Genes	At1g35680	Lsat_12741	TragDub11920-RA
Ribosomal Genes	At5g54600		TragDub11958-RA
Ribosomal Genes	At5g40950		TragDub18778-RA
Ribosomal Genes	At2g33450		TragDub26728-RA
Ribosomal Genes	At5g65220		TragDub13542-RA
Ribosomal Genes	At1g75350		TragDub09539-RA
Ribosomal Genes	At1g29070	Lsat_2139	TragDub29954-RA
Ribosomal Genes	At2g24090		TragDub28009-RA
Ribosomal Genes	At5g30510		TragDub17323-RA
Ribosomal Genes	At2g33800		TragDub07214-RA
Ribosomal Genes	At1g64510		TragDub15617-RA
Ribosomal Genes	At1g74970		TragDub18518-RA
Ribosomal Genes	At3g13120		TragDub28022-RA
Ribosomal Genes	At5g14320		TragDub09326-RA
Ribosomal Genes	At1g79850		TragDub10107-RA
Ribosomal Genes	At3g15190		TragDub09926-RA
Ribosomal Genes	At3g27160		TragDub28622-RA
Ribosomal Genes	At3g27830 and At3g27850[1]		TragDub12033-RA
Ribosomal Genes	At5g24490	Lsat_5980	
Ribosomal Genes	At1g68590	Lsat_3726	
Ribosomal Genes	At3g63490	Lsat_15799	
Ribosomal Genes	At2g43030	Lsat_1652	
Ribosomal Genes	At4g01310	Lsat_17875	

[1] *Paralogous proteins*

Table 2.4: List of identified putative proteins encoded by dual-targeted genes in *Tragopogon dubius.* The unique codes (e.g., At5g60960 and TragDub19304-RA) represent the associated genome loci of each protein in *Arabidopsis thaliana* and *T. dubius* genomes, respectively. Additional details of these proteins are given in Supplementary Table 2.1.

Cytonuclear Gene Category	*A. thaliana*	*L. sativa*	*T. dubius*
Dual-targeted Genes	At5g60960		TragDub19304-RA
Dual-targeted Genes	At1g14410		TragDub01137-RA
Dual-targeted Genes	At4g38160		TragDub06887-RA

In total, 21 control proteins that are not targeted to organellar genomes, involved in the OXPHOS alternative pathway, housekeeping genes or act as retrograde-signalling partners of dual-targeted genes, were identified (Table 2.5). These control proteins were identified to compare their expression patterns with those of organelle-targeted

cytonuclear genes (Chapter 3). Two *Arabidopsis* proteins in the OXPHOS alternative pathway (At3g22360 and At3g27620) could not be identified in both the *T. dubius* and *L. sativa* genomes.

Table 2.5: List of identified putative control proteins in *Tragopogon dubius* and *Lactuca sativa.* Red highlighted *T. dubius* protein was identified using the corresponding *L. sativa* protein. The unique codes (e.g., At4g28220, Last_31843 and TragDub21722-RA) represent the associated genome loci of each protein in *Arabidopsis thaliana*, *L. sativa* and *T. dubius* genomes, respectively. Additional details of these proteins are given in Supplementary Table 2.2.

Control Genes Category	***A. thaliana***	***L. sativa***	***T. dubius***
OXPHOS Alternative Pathway	At4g28220		TragDub21722-RA
OXPHOS Alternative Pathway	At4g21490	Lsat_31843	TragDub05309-RA
OXPHOS Alternative Pathway	At4g05020		TragDub18142-RA
OXPHOS Alternative Pathway	At2g20800		TragDub18142-RA
OXPHOS Alternative Pathway	At3g22370 and At1g32350[1]		TragDub14977-RA and TragDub14976-RA[1]
OXPHOS Alternative Pathway	At5g64210		TragDub08885-RA
OXPHOS Alternative Pathway	At1g07180		TragDub16181-RA
OXPHOS Alternative Pathway	At2g29990		TragDub16181-RA
OXPHOS Alternative Pathway	At5g08740	Lsat_36739	
Retrograde-signalling Genes	At4g26110		TragDub06418-RA and TragDub10196-RA[1]
Retrograde-signalling Genes	At1g58100		TragDub26682-RA
Retrograde-signalling Genes	At5g24680	Lsat_35148	
Retrograde-signalling Genes	At1g44920		TragDub21893-RA
Retrograde-signalling Genes	At5g01820		TragDub13849-RA and TragDub24073-RA[1]
Housekeeping Genes	At1g13440		TragDub11037-RA
Housekeeping Genes	At4g34270		TragDub09056-RA
Housekeeping Genes	At2g28390		TragDub05341-RA
Housekeeping Genes	At5g46630		TragDub22249-RA
Housekeeping Genes	At4g33380		TragDub05245-RA
Housekeeping Genes	At1g32750		TragDub07752-RA

[1] *Paralogous proteins*

2.4.2. Identification of Organellar (Plastid and Mitochondrial) Genes

The 66 organelle-encoded genes involved in selected plastid enzyme complexes and 22 genes implicated in mitochondrial OXPHOS complexes were retrieved from *T. dubius* plastid and *L. sativa* mitochondrial genomes, respectively (Table 2.6).

Table 2.6: List of identified organelle-encoded genes in *Tragopogon dubius* and *Lactuca sativa.*

Gene Name	Gene ID	Genome	Enzyme Complex
NAD(P)H dehydrogenase 1	nad1	Mitochondrial	OXPHOS Complex I
NADH dehydrogenase 2	nad2	Mitochondrial	OXPHOS Complex I
NADH dehydrogenase 3	nad3	Mitochondrial	OXPHOS Complex I
NADH dehydrogenase subunit 4	nad4	Mitochondrial	OXPHOS Complex I
NADH dehydrogenase subunit 4L	nad4L	Mitochondrial	OXPHOS Complex I
NADH dehydrogenase subunit 5	nad5	Mitochondrial	OXPHOS Complex I
NADH dehydrogenase 6	nad6	Mitochondrial	OXPHOS Complex I
NADH dehydrogenase subunit 7	nad7	Mitochondrial	OXPHOS Complex I
NADH dehydrogenase subunit 9	nad9	Mitochondrial	OXPHOS Complex I
Succinate dehydrogenase 3-1	sdh3a	Mitochondrial	OXPHOS Complex II
Succinate dehydrogenase 3-2	sdh3b	Mitochondrial	OXPHOS Complex II
Succinate dehydrogenase subunit 4	sdh4	Mitochondrial	OXPHOS Complex II
Apocytochrome B	cob	Mitochondrial	OXPHOS Complex III
Cytochrome oxidase subunit 1	cox1	Mitochondrial	OXPHOS Complex IV
Cytochrome oxidase subunit 2	cox2	Mitochondrial	OXPHOS Complex IV
Cytochrome C oxidase subunit 3	cox3	Mitochondrial	OXPHOS Complex IV
ATP synthase subunit 1-A	atp1A	Mitochondrial	OXPHOS Complex V
ATP synthase subunit 1-B	atp1B	Mitochondrial	OXPHOS Complex V
ORF25	atp4	Mitochondrial	OXPHOS Complex V
ATPase subunit 6	atp6	Mitochondrial	OXPHOS Complex V
ATP synthase 8	atp8	Mitochondrial	OXPHOS Complex V
Mitochondrial F0-ATPase subunit 9	atp9	Mitochondrial	OXPHOS Complex V
PSAA	psaA	Plastid	Photosystem I
PSAB	psaB	Plastid	Photosystem I
PSAC	psaC	Plastid	Photosystem I
Photsystem I subunit I	psaI	Plastid	Photosystem I
PSAJ	psaJ	Plastid	Photosystem I
Photosystem II reaction centre protein A	psbA	Plastid	Photosystem II
Photosystem II reaction centre protein B	psbB	Plastid	Photosystem II
Photosystem II reaction centre protein C	psbC	Plastid	Photosystem II
Photosystem II reaction centre protein D	psbD	Plastid	Photosystem II
Photosystem II reaction centre protein E	psbE	Plastid	Photosystem II
Photosystem II reaction centre protein F	psbF	Plastid	Photosystem II
Photosystem II reaction centre protein H	psbH	Plastid	Photosystem II
Photosystem II reaction centre protein I	psbI	Plastid	Photosystem II
Photosystem II reaction centre protein J	psbJ	Plastid	Photosystem II
Photosystem II reaction centre protein K precursor	psbK	Plastid	Photosystem II
Photosystem II reaction centre protein L	psbL	Plastid	Photosystem II
Photosystem II reaction centre protein M	psbM	Plastid	Photosystem II
Photosystem II reaction centre protein N	psbN	Plastid	Photosystem II
Photosystem II reaction centre protein T	psbT	Plastid	Photosystem II
YCF9	psbZ	Plastid	Photosystem II
RNA polymerase subunit alpha	rpoA	Plastid	RNA polymerase PEP
RNA polymerase subunit beta	rpoB	Plastid	RNA polymerase PEP
RPOC1	rpoC1	Plastid	RNA polymerase PEP

Gene Name	Gene ID	Genome	Enzyme Complex
RPOC2	rpoC2	Plastid	RNA polymerase PEP
ATP synthase subunit alpha	atpA	Plastid	ATP Synthase
ATP synthase subunit beta	atpB	Plastid	ATP Synthase
ATP synthase epsilon chain	atpE	Plastid	ATP Synthase
ATPF	atpF	Plastid	ATP Synthase
ATPH	atpH	Plastid	ATP Synthase
ATPI	atpI	Plastid	ATP Synthase
Photosynthetic electron transfer A	petA	Plastid	Cytochrome b6f
Photosynthetic electron transfer B	petB	Plastid	Cytochrome b6f
Photosynthetic electron transfer D	petD	Plastid	Cytochrome b6f
PETG	petG	Plastid	Cytochrome b6f
PETL/ORF31	petL	Plastid	Cytochrome b6f
PETN/YCF6	petN	Plastid	Cytochrome b6f
Caseinolytic protease P 1/Plastid-encoded CLP P	clpP	Plastid	Caseinolytic Proteases
Accd, acetyl-coa carboxylase carboxyl transferase subunit beta	accD	Plastid	Acetyl-CoA Carboxylase
RBCL	rbcL	Plastid	Rubisco
Ribosomal protein L2	rpl2	Plastid	Ribosomal Genes
Ribosomal protein L14	rpl14	Plastid	Ribosomal Genes
Ribosomal protein L16	rpl16	Plastid	Ribosomal Genes
Ribosomal protein L20	rpl20	Plastid	Ribosomal Genes
Ribosomal protein L22	rpl22	Plastid	Ribosomal Genes
Ribosomal protein L23	rpl23	Plastid	Ribosomal Genes
Ribosomal protein L32	rpl32	Plastid	Ribosomal Genes
Ribosomal protein L33	rpl33	Plastid	Ribosomal Genes
Ribosomal protein L36	rpl36	Plastid	Ribosomal Genes
Ribosomal protein S2	rps2	Plastid	Ribosomal Genes
Ribosomal protein S3	rps3	Plastid	Ribosomal Genes
Chloroplast ribosomal protein S4	rps4	Plastid	Ribosomal Genes
Chloroplast ribosomal protein S7	rps7	Plastid	Ribosomal Genes
Ribosomal protein S8	rps8	Plastid	Ribosomal Genes
Ribosomal protein S11	rps11	Plastid	Ribosomal Genes
Ribosomal protein S12A	rps12a	Plastid	Ribosomal Genes
Ribosomal protein S12B	rps12b	Plastid	Ribosomal Genes
Chloroplast ribosomal protein S14	rps14	Plastid	Ribosomal Genes
Chloroplast ribosomal protein S15	rps15	Plastid	Ribosomal Genes
Ribosomal protein S16	rps16	Plastid	Ribosomal Genes
Ribosomal protein S18	rps18	Plastid	Ribosomal Genes
Ribosomal protein S19	rps19	Plastid	Ribosomal Genes

2.5. Discussion

Cytonuclear coordination and co-evolution are crucial mechanisms for the effective regulation of plant cells (Greiner and Bock 2013). Cytonuclear enzyme complexes are the hub of cytonuclear interactions and are an excellent model to investigate the evolutionary consequences of cytonuclear integration and coordination. In addition to these interacting multi-subunit enzyme complexes, other nuclear-encoded proteins also interact with the organellar genomes. For example, dual-targeted genes interact and have dual functional roles, including DNA replication/recombination, DNA repair, post-transcriptional modifications and tRNA aminoacylation etc., in both plastid and mitochondria (Carrie and Small 2013; Forsythe et al. 2019). In the current study, cytonuclear genes as well as dual-targeted genes were identified in a draft *T. dubius* genome to study the cytonuclear co-evolution in *Tragopogon* allopolyploid species.

In flowering plants, the endosymbiotic transfer of organellar genes to the nucleus is an extremely active and continual process compared to other eukaryotes which have achieved long-term stability in their cytoplasmic gene content (Timmis et al. 2004; Johnston and Williams 2016). This may result in the altered cytonuclear interactions and integration across plant species. For example, the OXPHOS complex II is entirely nuclear-encoded in *Arabidopsis thaliana* (Adams et al. 2002; Forsythe et al. 2019). However, complex II subunits *sdh3* and *sdh4* have been retained in the mitochondrial genomes of several other angiosperm species, including *L. sativa* (Adams et al. 2001). In the current study, the putative orthologs of the *Arabidopsis* OXPHOS complex II genes *sdh3* and *sdh4* were not found in the *T. dubius* draft genome. This could suggest that these genes in *Tragopogon* species might also be preserved in the mitochondrial genome, making the OXPHOS complex II a cytonuclear complex in *Tragopogon* species.

However, this would need to be confirmed through the analysis of the completely annotated nuclear and mitochondrial genomes of *Tragopogon* species.

The gene composition of cytonuclear enzyme complexes is lineage-specific, meaning that it differs across different linages. For example, the OXPHOS complex I is comprised of more than 40 subunits, e.g., 45 in mammals (Carroll et al. 2006), 48 in *Arabidopsis* (Klodmann et al. 2010) and 42 in the green alga, *Chlamydomonas reinhardtii* (Sazanov 2015; Cardol et al. 2004). In this study, we have identified 38 putative protein subunits for the OXPHOS complex I in *T. dubius* species and for the remaining five missing *Arabidopsis* proteins, we have found only 4 putative protein orthologs in the related species, *L. sativa*. The one protein (gamma carbonic anhydrase 3) that we could not find even in the *Lactuca* genome is an accessory subunit of complex I. There are five gamma carbonic anhydrase subunits (*CA1*, *CA2*, *CA3*, *CAL1* and *CAL2*) present in *Arabidopsis* (Parisi et al. 2004; Perales et al. 2004). However, in *T. dubius*, two gamma carbonic anhydrases, TragDub22387-RA and TragDub01208-RA, were retrieved against *CA1* & *CA2* and *CAL1* & *CAL2*, respectively. Similarly, in other plant species including rice, poplar and maize, only one gene encoding *CAL* has been found (Meyer 2012). In the case of other anhydrases (*CA1*, *CA2* and *CA3*), plant species either have all three (e.g., *Arabidopsis*, poplar and rice) or only two of them (*CA1* and *CA2*) as observed in sorghum and maize (Meyer 2012). Based on these findings, we suggest that the OXPHOS complex I in *T. dubius* might have only two carbonic anhydrase subunits, which may contribute to the differential subunit composition in *T. dubius* compared to *Arabidopsis* and other plant species.

The number of identified putative protein orthologs involved in other selected mito-nuclear and plastid-nuclear interacting enzyme complexes in *T. dubius* also differed in the subunit composition compared to *A. thaliana* (Table 2.1). The heterogeneity in the

gene content resulted from the fact that for certain *Arabidopsis* gene paralogs, there was only a single-copy ortholog found in *T. dubius*. This could suggest that these genes in *Tragopogon* have not undergone gene duplication events. For example, for paralogs of some *Arabidopsis* genes, including NADH-ubiquinone oxidoreductase-related gene and Alpha-helical ferredoxin gene in OXPHOS complex I, photosystem I subunits D and H as well as photosystem II subunit Q, single-copy orthologs were identified in the *T. dubius* genome (Tables 2.2 & 2.3). Additionally, a few *Arabidopsis* genes belonging to various enzyme complexes were not found at all in the *T. dubius* draft genome (Supplementary Table 2.3). For instance, three *Arabidopsis* genes from each cytonuclear complex including ribosomal genes, photosystem II, OXPHOS complex III and IV were missing in both the *T. dubius* and *L. sativa* genomes (Supplementary Table 2.3). Because a draft *T. dubius* genome was used for gene screening, it could be possible these genes were missing in the draft genome annotation, or they might not exist at all in the *Tragopogon* species, as we were not able to identify these genes even in the related species, *L. sativa* genome. Lastly, it could be possible that these *Tragopogon* genes are divergent and have no substantial similarity to the *Arabidopsis* genes, considering *A. thaliana* is not a close relative to *T. dubius* and belongs to a separate family (Brassicaceae) of plants. In future, gene identification using updated versions of the *Tragopogon* species genomes would be a better way to verify these 'missing' genes and to further curate a complete list of genes involved in different genetic pathways and networks.

Chapter 3

Gene Expression Analysis of Duplicated Nuclear Genes (Homeologs) Implicated in Cytonuclear Interactions in *T. miscellus* Allopolyploids

3.1. Abstract

Allopolyploidy is considered one of the major pathways for plant species diversification. In allopolyploids, the duplicated and biparentally inherited nuclear genomes interact with only one set of maternally inherited cytoplasmic (mitochondrial and plastid) genomes, which results in an altered cytonuclear stoichiometry. In this project, we employed RNA CaptureSeq to examine expression differences of cytonuclear genes in *Tragopogon miscellus* allopolyploids along with their diploid parental species, *Tragopogon dubius* and *Tragopogon pratensis*, at different developmental stages. These are young, naturally occurring allopolyploids, which have formed multiple times, including reciprocally, hence they offer a window into the potential for repeated evolution of cytonuclear coordination. Using a *T. dubius* draft genome, we developed probes for ~293 loci involved in nuclear-mitochondrial and nuclear-plastid complexes as well as dual-targeted genes. In short- and long-liguled *T. miscellus*, ~90-98% duplicated nuclear genes/homeologs showed no difference in the total expression level compared to their parent species, at all growth stages. And only 1-3% genes generally displayed unbalanced expression level dominance toward their respective paternal parent in both reciprocal forms. In terms of homeolog expression bias, the homeologs with biased expression in short- (52%) and long-liguled (43%) *T. miscellus* allopolyploids generally showed unbalanced homeolog expression bias toward *T. pratensis* across all developmental stages. In comparison to the parental

species, 42-56% and 38-41% of the loci in short- and long-liguled *T. miscellus*, respectively, exhibited novel homeolog expression patterns. The genes that showed either diverse patterns of homeolog expression bias or expression level dominance were determined to belong to different cytonuclear complexes rather than being enriched to a certain complex. The existence of such flexible and diverse expression patterns in *T. miscellus* allopolyploids depicts the complex role of allopolyploidization in co-regulating the cytonuclear genomes.

3.2. Introduction

Allopolyploidy, which involves interspecific hybridization and whole genome duplication, is considered a major driving force in the evolution of plant species (Soltis et al. 2016; Madlung 2013b; Van de Peer et al. 2017b). Ancient whole genome duplication events are prevalent in almost all angiosperms (Jiao et al. 2011a) and 25-35% of plant species have also encountered recent polyploidy events (Salman-Minkov et al. 2016a). The high prevalence of polyploidization events in plants may assist polyploids against environmental stresses (Wu et al. 2020), in gaining novel traits (Soltis et al. 2009; Soltis and Soltis 2016) and faster niche differentiation compared to their diploid parents (Baniaga et al. 2020). To date, studies conducted on different allopolyploids and their diploid ancestors have demonstrated that whole genome duplication results in various complex genomic changes, including alterations in epigenetic modifications, gene loss, expression level dominance/transgressive expression (non-additive gene expression), homeolog expression bias, additive expression and changes at the chromosomal level (Figure 3.1) (reviewed in Osborn et al. 2003b; Doyle et al. 2008; Buggs et al. 2014; Doyle and Coate 2019; Wendel et al. 2018a; Madlung and Wendel 2013; Yoo et al. 2014a). In allopolyploids, expression level dominance and homeolog expression bias can either be balanced (equal number of loci show expression level dominance/homeolog expression bias toward each sub-genome) or unbalanced (more genes show expression level dominance/homeolog expression bias toward one sub-genome over the other). In addition to the aforementioned changes, the merger of duplicated and biparentally inherited diverged nuclear genomes in allopolyploids along with only one set of maternally inherited cytoplasmic (mitochondrial and plastid) genomes, also brings together complex interactions between these tripartite genomes. This ultimately may result in cytonuclear

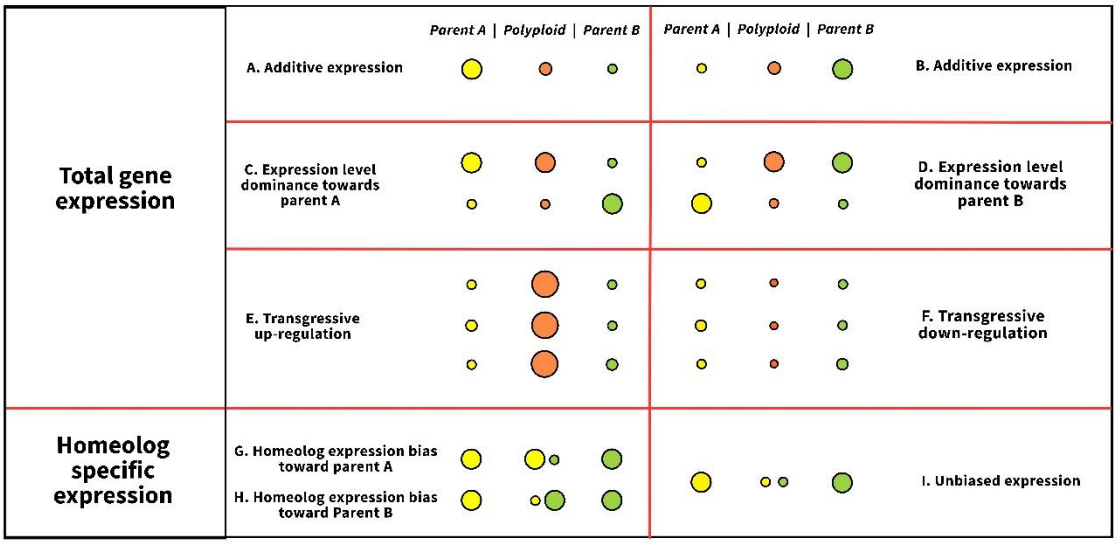


Figure 3.1: Graphical representation of additive and non-additive expression patterns in allopolyploids for total gene expression and homeolog-specific expression. Circles depict the expression levels of a particular gene, and the size of the circles represents the relative expression level of a gene. A & B) Additive expression occurs when the total gene expression in an allopolyploid is additive (i.e., the sum) of the two parents, who might differ in their expression profiles (lower in one parent and higher in the other). C & D) Expression level dominance occurs when the total expression of duplicated genes is statistically similar to one of the diploid parents. For example, scenarios C and D show expression level dominance toward parent A or B, respectively. E & F) Transgressive expression patterns occur when the gene expression levels in a polyploid are either up- (E) or down-regulated (F) compared to both parents. G & H) Homeolog expression bias is the relative contribution of the homeologs to the total gene expression. This can be biased toward either parental species or unbiased (I).

incompatibilities in allopolyploids due to divergence and evolutionary differences among these genomes, which can have profound effects on a plant's fitness (Fishman and Willis 2006; Hill 2017; Postel and Touzet 2020; Fishman and Sweigart 2018). For instance, interactions between nuclear- and cytoplasmic-encoded subunits of chimeric enzyme complexes (discussed in Chapter 1; section 1.7) play central roles in eukaryotic bioenergetics as these enzyme complexes are involved in photosynthesis, organelle ribosomes and cellular respiration (Rand et al. 2004). Also, nuclear-encoded proteins mediate DNA replication/repair, transcriptional and post-transcriptional processes of organelle genomes through direct interaction with either DNA or their RNAs (Salinas-

Giegé et al. 2015; Gualberto and Newton 2017; Germain et al. 2013; Zhang et al. 2016). Hence, the dosage imbalances between these genomes may cause potential disruption of the function and assembly of chimeric complexes. Consequently, this poses a major challenge for allopolyploids with potential immediate consequences including altered cytonuclear stoichiometry, which in turn can impact important physiological processes such as photosynthesis and cellular respiration.

To cope with challenges associated with cytonuclear incompatibilities following whole genome duplications, allopolyploids may undergo alterations to synchronize imbalances between nuclear and cytoplasmic genomes. These mechanisms might include up-regulation of organelle genes and their copy number, up-regulation of maternal homeologs with down-regulation of paternal homeologs, maternal gene conversion events through homeologous exchanges and/or loss of function of paternal homeologs (pseudogenization) (Doyle and Coate 2019; Sharbrough et al. 2017).

Although cytonuclear co-evolution has long been recognized as a potential evolutionary barrier to speciation, these cytonuclear interactions have not been widely explored in allopolyploids. Thus far, these interactions have mostly been investigated with reference to plastid–nuclear incompatibilities with a focus on the Rubisco enzyme complex (Ribulose 1, 5-bisphosphate carboxylase/oxygenase). For example, Gong et al. (2012) reported maternal to paternal gene conversion events between nuclear homeologs through non-reciprocal homeologous recombination. Moreover, preferential expression of maternal homeologs was observed in both cultivated and wild cotton allopolyploids (Gong et al. 2012) and later between homeologous genomes of *Arachis hypogaea*, *Nicotiana tabacum* and *Arabidopsis suecica* allopolyploids (Gong et al. 2014). However, studies conducted in a resynthesized allotetraploid of rice, *Oryza sativa*, or in young allopolyploids of *Tragopogon miscellus* did not report homogenization of nuclear

homeologs through gene conversion events (Sehrish et al. 2015; Wang et al. 2017a). Additionally, synthesized rice allotetraploids showed no specific pattern of transcriptional bias toward the maternal parent (Wang et al. 2017a). Furthermore, Ferreira de Carvalho et al. (2019) reported dominance of nuclear genes involved in various plastid complexes towards the paternal sub-genome and found no evidence of homogenization via maternal gene conversions in resynthesized and natural *B. napus* allopolyploids. Likewise, a study in a synthesized *Cucumis* allopolyploid revealed paternally biased expression of nuclear *rbcS* gene (Zhai et al. 2019b). Lastly, Grover et al. (2022) recently conducted an extensive survey of a large number of genes targeted to cytoplasmic genomes as well as genes that were not targeted to organelles in multiple allopolyploid systems. They noted varying patterns of cytonuclear accommodation among lineages, cytonuclear gene categories, as well as in genes within a particular category. Also, this study did not find strong evidence of maternal bias to balance cytonuclear interactions in the examined allopolyploids. Taking into consideration this vast array of outcomes, one can say that maternally biased transcription of the nuclear homeologs encoding the cytonuclear enzyme complexes is not always the rule in allopolyploids, as we might expect. Instead, diverse responses to accommodate the interaction among the genomes are typical. Therefore, to understand the mechanisms that stabilize cytonuclear interactions, here we investigated the gene expression of nuclear homeologs, and their organelle counterparts involved in mitochondrial oxidative phosphorylation, plastid complexes, as well as dual-targeted genes in *Tragopogon miscellus* allopolyploids. *T. miscellus* allopolyploids offer a unique system to study recurrent and recent (90-100 years old) allopolyploidy in natural populations (Soltis et al. 2004). Repeated formation of *Tragopogon* allopolyploids, including short- and long-liguled reciprocal forms (Soltis et al. 1995a; Soltis et al. 2004; Symonds et al. 2010) provides a way to determine if such independently formed

allopolyploid populations develop similar cytonuclear coordination. Interestingly, only one natural population of the long-liguled form but multiple populations of the short liguled form exist in nature, as discussed in chapter 1 (section 1.9). RNA CaptureSeq (Targeted RNAseq) was used to evaluate the selected transcripts in plants from natural populations of the reciprocally formed *T. miscellus* allopolyploids and progenitor (*T. pratensis* and *T. dubius*) species. The expression patterns among all species were examined to test the following hypotheses:

i. Duplicated nuclear genes involved in cytonuclear enzyme complexes as well as dual-targeted genes will exhibit maternally biased expression.

ii. Reciprocal forms of *T. miscellus* will show biased homeolog expression, toward their respective maternal parent.

iii. Cytonuclear genes will undergo development-mediated differential regulation of homeologs.

iv. To compensate for the duplicated nuclear genomes, the expression of maternally inherited organelle genes will be up-regulated in reciprocally formed allopolyploids.

3.3. Materials and Methods

3.3.1. Plant Materials

Seeds of diploid parental species, *Tragopogon dubius* and *T. pratensis*, as well as short- and long-liguled *T. miscellus* allopolyploids were collected from natural populations in the United States (2016 and 2017) (Figure 3.2) and were grown in a plant growth chamber at Massey University under controlled conditions (20˚C and 16/8 hours light/dark). Seventeen individuals including 1-3 replicates per lineage (individual) from five natural populations of the two diploid parents were included in this study. For *T.*

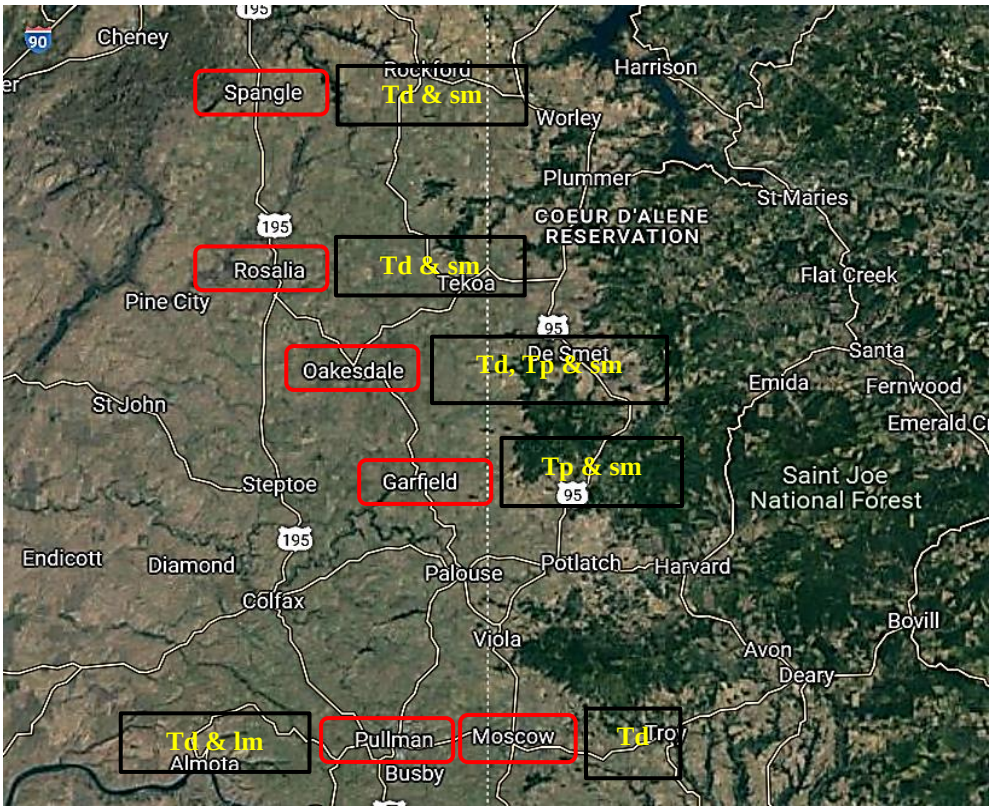


Figure 3.2: The map shows USA collection sites (encircled red) where seed collections of natural populations of both parental species and reciprocal forms of *Tragopogon miscellus* allopolyploids were made for this study. The markings besides each location represent the species collected at a particular site. (Td: *T. dubius*, Tp: *T pratensis*, sm: short-liguled *T. miscellus* and lm: long-liguled *T. miscellus*).

miscellus, 21 individuals (2-3 replicates per lineage) from five populations were included. In total 76 plants (Supplementary Table 3.1) were grown in 4" plastic pots with Tui Pot Power soil (Tui, New Zealand) and later were transplanted to larger containers (1.4 L) after five weeks of germination. As some plants flowered, these were transplanted into bigger pots (7 L) with fresh media to allow further growth. A Randomized Block Design was followed to randomly allocate pots into each block (tray) so that each block had a mix of different lineages of parents and allopolyploid species. In total, there were 7 trays (blocks) with 12 samples per tray. Trays were also rotated each week to avoid any positioning effects on plant's growth.

3.3.2. Sample Collection

Leaf tissues (4 cm, youngest leaf from base to tip) were collected at young (9 weeks after germination, hereafter referred to as the early-age) and middle-age (24 weeks post-germination) growth stages directly into liquid nitrogen and then stored at -80°C until further processing. *Tragopogon* normally requires cold-treatment to induce flowering, however some individuals of *T. dubius* (5 plants), *T. pratensis* (1 plant) and short-liguled *T. miscellus* (6 plants) started flowering without a cold treatment. These individuals are referred to as the 'early-flowering group' and are represented by young (9 weeks) and mature-age growth stages only. For the latter, cauline leaves from these early flowering lineages were collected 22 weeks following germination. The other plants (52 plants) that did not flower early are referred to as the 'late-flowering group' and for these, leaf tissue was collected at the early (9 weeks) and middle-growth (24 weeks) stages.

3.3.3. RNA Extraction, Quantification and Quality Check

RNA extractions from leaf tissues of 76 samples (Supplementary Table 3.1), including all candidate species, were performed using the RNeasy Plant Mini Kit (Qiagen), including an on-column DNase digestion using the RNase free DNase set (Qiagen) as per kit's instructions. RNA was quantified using a Qubit Fluorometer (Invitrogen™) and purity confirmed by measuring A260/A280=2 and A260/A230=2-2.2 ratios via a Nanondrop spectrophotometer (Thermo Fisher Scientific™ NanoDrop™ One/OneC Microvolume UV-Vis Spectrophotometer). The RNA integrity was assessed by running RNA samples on a LabChip® GX Touch HT (PerkinElmer, United States) using the Standard LabChip® assay at Massey Genome Service (Palmerston North, New Zealand). RNA samples with RQS (RNA Quality Score) 6.5 and above were later used for RNA CaptureSeq library preparation.

3.3.4. RNA CaptureSeq

RNA CaptureSeq is a targeted RNASeq that involves the construction of oligonucleotide probes against selected genes of interest, followed by their hybridization to cDNA libraries and sequencing (Mercer et al. 2012). This approach was used to target 316 genes (discussed in chapter 2, Tables 2.2-2.5 & Supplementary Table 2.3) via custom probes designed and manufactured by Roche Nimblegen, to analyse their expression in both allopolyploid forms and parental species. RNA CaptureSeq procedures are described in the following sections and a brief overview of the workflow is shown in the Figure 3.3.

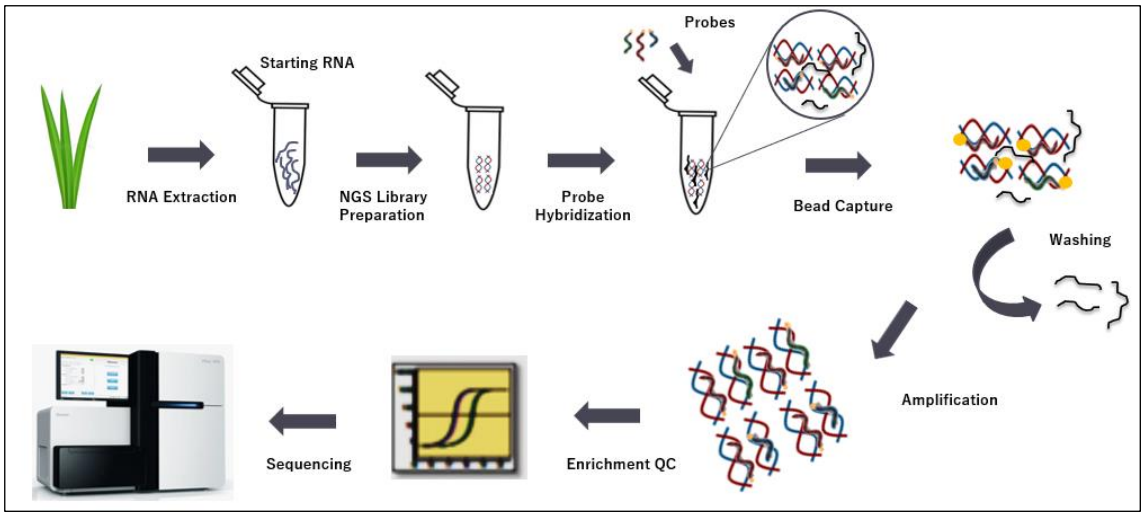


Figure 3.3: Overview of RNA CaptureSeq workflow. RNA extracted from leaf tissues of each individual was used as starting material for the pre-captured library preparation. Following the library preparation, all the libraries were pooled together to carry out hybridization using custom designed probes. After probe hybridization the probe-hybridized transcripts were captured using bead capture followed by washing. Next amplification of the captured library pool was done, followed by QC enrichment using qPCR. The quality-assured library pool was sequenced on Illumina HiSeq.

3.3.4.1. RNASeq Pre-captured Library Preparation, Quantification and Quality Assessment

Following RNA isolation, RNASeq pre-captured libraries were prepared using KAPA mRNA HyperPrep Kit (Roche) for Illumina® platforms and each library was uniquely indexed using adapters available in KAPA Unique Dual-Indexed Adapter Kit (15 µM; Roche). Initially, library preps for 33 samples used 250 ng of total RNA per

sample. Also, 2 µl of ERCC RNA spike-in mixes 1 and 2, included in the ERCC ExFold RNA Spike-In Mixes kit (Invitrogen™), were added separately to each total RNA sample at 1:400 dilution, according to the user guidelines. After library preparation, these final pre-captured libraries were quantified via Qubit assay. The sizing profile of library fragments and adapter dimer contamination were determined by running libraries on a LabChip® GX Touch HT using the DNA High Sensitivity LabChip® assay at Massey Genome Service (Palmerston North, New Zealand). Most of these pre-captured libraries showed adapter dimers, as indicated by a sharp 120-140 bp peak. Therefore, a second 1X bead clean-up post-amplification was performed to remove adapter dimers as per kit's recommendation. Additionally, to avoid adapter dimers in the remaining library preps, the total amount of input RNA was increased to 500-1000 ng. Based on the increased amount of RNA, the adapter ligation step of the protocol was modified to use diluted adapter stock (5 µl) in 1.5 µM concentration instead of 7 µM (recommended concentration for 500-1000 ng input RNA as per kit's protocol). In the case of two samples, where insufficient RNA remained, 100 ng and 250 ng of total RNA was used for the library preps. Thus, to compensate for the quantity of starting material and to avoid adapter dimers in these libraries, 2 and 3 µl of 1.5 µM diluted adapter stock were added, respectively, instead of the recommended 5 µl. PCR grade water was used to make up the remaining diluted adapter stock volume. Likewise, 1 µl of ERCC spike-in mixes was added into all remaining RNA samples at 1:100 dilution based on the RNA amount (100 ng, 500 ng & 1000 ng) used. All final pre-captured libraries were quantified, and the sizing profile of the library fragments and adapter dimer contamination were assessed as previously described.

3.3.4.2. RNASeq Captured Multiplex Library Preparation, Quantification and Quality Assessment and RNA Sequencing

A captured multiplex library was prepared using the SeqCap EZ probe pool, which included probes for the 316 genes identified and discussed in the previous chapter as well as probes for capturing the ERCC spike-ins (included in the probe set for later validation of the RNASeq platform), the HyperCap Target Enrichment Kit (Roche) and the HyperCap Bead Kit (Roche). At first, 76 pre-captured libraries made in the previous step were pooled together in equal amounts (by mass) to obtain a single multiplex library pool of combined mass of 1.25 µg. After pooling the libraries together, the concentration of the pre-captured library pool was determined via Qubit. One µg of this library pool was further used for the probe hybridization, followed by recovery of the final captured multiplex library using the kits mentioned above, following kits' instructions. Next, the captured multiplex library was quantified using Qubit. As an internal quality check to confirm the success of probe hybridization and to assess the level of enrichment achieved from captured multiplex library, a qPCR enrichment assay using LightCycler® 480 SYBR Green I 2X Master Mix was accomplished on LightCycler® 480 Instrument II (Roche). This qPCR was done using four templates: the pre-captured library pool, the captured multiplex library pool, cDNA of any two parental samples included in the library pool as positive controls, and a negative control. Three housekeeping genes targeted through this assay were included in the probe design (Chapter 2, Table 2.5), thus, providing an effective way to determine the success of targets' capture.

Prior to next-generation sequencing, the quantification and sizing profile of the library pool using the DNA High Sensitivity LabChip® assay were checked, as well as a qPCR assay to confirm the quality of the captured pool. The high-quality captured multiplex library was then submitted to Novogene Co., Ltd. (Singapore) for paired-end

sequencing (150 bp reads) on an Illumina HiSeq system.

3.3.5. RNA CaptureSeq Data Analysis

3.3.5.1. Qualitative Analysis of Raw Reads

The preliminary quality check of sequencing data was done by Novogene Co., Ltd. (Singapore). This quality check included filtering of raw reads to remove reads with low quality bases (Q score < 5) that constitute more than 50% of the total bases and reads containing N > 10%. This filtering process also removed reads containing the sequencing adaptors. After receiving the processed data from Novogene, an additional qualitative analysis, including pre-processing of raw reads, was done to trim adapters and remove low-quality bases using fastp (version 0.20.0) (Chen et al. 2018).

3.3.5.2. Mapping and Classification of Raw Reads with respect to Homeologs

Filtered and good quality raw reads were mapped to the reference gene models, including ERCC transcript sequences, using STAR (version 2.7.7a) (Dobin et al. 2012) and the mapping statistics evaluated to validate our probe/capture design. Next, mapped reads were analysed using HyLiTE (Duchemin et al. 2015), which classified reads to each homeolog (parental) class based on SNPs, and also generated a comprehensive file with expression levels for each gene in the allopolyploids and diploid parent species. These files with information about homeolog read counts, as well as total read counts per gene, were used in the downstream analyses.

3.3.5.3. Quality Validation of RNASeq Platform

To evaluate the RNA CaptureSeq platform in terms of technical and biological signal, two types of analyses were carried out: ERCC-based analysis, and hierarchical

clustering and Principal Components Analysis (PCA) as described below.

ERCC RNAs Based Analysis

This analysis included assessment of the RNASeq platform with respect to the dynamic range and fold-change response. For the former analysis, a plot was made for each sample using DESeq2 (version 1.32.0) (Love et al. 2014) with normalised read counts of each ERCC transcript against their known molar concentration, which was added in the input total RNA at the start of library preparation. A linear regression model was used to determine the relationship between input ERCC transcripts' concentration and their output signal through the best fit line.

For the fold-change response analysis, the observed fold change ratio of Mix 1:Mix 2 for each ERCC transcript was compared with their expected fold-change ratio and their correlation was determined by a linear regression model. To identify genes with differential expression among samples with different spike-in mixes, the observed fold-change ratio for each ERCC transcript was obtained by performing differential expression analysis via DESeq2 (version 1.32.0) (Love et al. 2014) using default parameter settings. For this differential expression analysis, raw read counts mapped to each ERCC transcript, as well as candidate genes across all samples, were used and normalised with ERCC spike-ins specified in the design formula of the DESeq2 dataset object.

Hierarchical Clustering and Principal Component Analysis

To assess if the differences in the sequencing data were biologically accurate and to identify any sample outliers, principal component analysis and hierarchical clustering were performed at the sample level. For both analyses, rlog transformed read count data were used to make a PCA chart as well as clustered heatmaps.

3.3.5.4. Differential Gene Expression Analysis

Total gene expression levels were examined for differential expression among both parental species and reciprocal forms of *T. miscellus* at each growth stage. In the case of early growth stage samples, three types of differential expression analyses were carried out: one including all early-stage samples of both early- and late-flowering plant groups and for the other two analyses, early-stage samples of each group were considered independently. These differential gene expression analyses were carried out using DESeq2 (version 1.32.0) with the negative binomial generalized linear model (Love et al. 2014). Raw read counts of all genes, except ERCC spike-in transcripts and the *Arabidopsis*-derived genes (which did not successfully target the *Tragopogon* species), were used for this purpose. Read counts were normalised through default normalisation settings in the DESeq2 package, which takes sequencing depth and RNA composition into account. At the gene level, genes with 0 read counts across all samples and low mean normalised counts (<10) were filtered out from the data (Shan et al. 2020a). A gene was considered to be significantly differentially expressed if the p-adjusted value was less than 0.05 and the $\log_2$(fold change) was greater than 1.5.

3.3.5.5. Total Gene Expression Levels

Differentially expressed genes identified by the total gene expression levels were partitioned into distinct expression patterns including additive and non-additive gene expression (Rapp et al. 2009b; Shan et al. 2020a; Grover et al. 2012b). Non-additive expression patterns were further subdivided into expression level dominance and transgressive expression (discussed in detail in Chapter 1, section 1.5.2). Briefly, a gene displayed additive expression when its total expression level in the allopolyploid was higher than one of the parent species but lower than the other parental species, considering

that gene is being expressed differentially between parent species (Figure 3.1, categories A & B) (Grover et al. 2012b). Expression level dominance was determined when the total expression level of a gene in the allopolyploid was comparable to one of the parent species (Figure 3.1, categories C & D); transgressive expression was noted when the gene expression level in the allopolyploid was either higher or lower than the expression of that gene in diploid parent species (Figure 3.1, categories E & F) (Grover et al. 2012b).

3.3.5.6. Homeolog Specific Analyses

Reads assigned to each homeolog via HyLiTE were also analysed for differential expression using DESeq2 (version 1.32.0) to determine the relative contribution of homeolog expression in the reciprocal forms of *T. miscellus*. Homeologs were then categorized into the following patterns: homeolog expression biased toward either parent species (Figure 3.1, categories G & H) or unbiased expression (Figure 3.1, category I). Moreover, in comparison to expression differences of loci between diploid parent species, the homeologs in the allopolyploids were also grouped into three characteristic expression categories: parental legacy (homeolog expression patterns in an allopolyploid are similar to those in the diploid parents), no homeolog expression bias or novel homeolog expression bias.

3.3.5.7. Organelle Gene Expression

Because organelle genes encoded by the mitochondrial and plastid genomes are usually inherited from the maternal parent in *Tragopogon* allopolyploids (Sehrish et al. 2015) and did not undergo duplication via an allopolyploidization event, their overall general expression patterns were examined via hierarchical clustering analysis. Also, to identify differentially expressed organelle genes, three types of differential expression

analyses were performed using DESeq2: between parent species, between each population of short- and long-liguled *T. miscellus*, and between reciprocal allopolyploid forms and their respective maternal parents.

3.4. Results

3.4.1. Mapping Statistics Support Probe Capture Design

After quality check of raw reads, a promising percentage (78-84%) of reads uniquely mapped to the reference gene models in all samples including multiple individuals from various natural populations of *T. dubius*, *T. pratensis* and reciprocal forms of *T. miscellus* allopolyploids (Supplementary Figure 3.1). Additionally, the ratio of multi-mapping reads that mapped to multiple gene models was low for each individual sample, ranging between 5-10% (Supplementary Figure 3.2). Likewise, the mean proportion of mismatches between sequencing reads and the reference gene models was also very low (0.36-1.42%) (Supplementary Figure 3.3). *T. dubius* samples had an expected low mismatch rate across the entire sample set, because probes were designed against candidate gene sequences retrieved from a *T. dubius* draft genome. Although the CaptureSeq probes were designed against only the *T. dubius* parent, one might expect a higher proportion of mismatches between sequencing reads and the other parental species, *T. pratensis*. However, mismatch rates were still quite low for *T. pratensis* (1.42%), considering the probes were designed against the *T. dubius* draft genome. This may be because the genomes of both parent species are highly similar and high-density probes were able to capture genes from both genomes efficiently. All these mapping statistics suggested that the capture design uniquely captured specific copies of gene paralogs in all species and successfully enriched RNASeq libraries for genes of interest, suggesting that our capture design worked well for this project.

3.4.2. Validation of RNA CaptureSeq platform

ERCC RNAs based Analysis

Our RNASeq libraries included a set of external controls (ERCC-spike in transcripts) with known molarity that allowed us to test the platform's ability to detect transcripts across a range of input concentrations and the sensitivity of differential expression analyses. Two types of analyses (dose response and fold change response) were done in this regard. In terms of dynamic range and dose response, a linear relationship between the reads aligned to 92 ERCC spike-in transcripts and their known molar concentration was observed (Figure 3.4). A separate plot was made for each sample library, and similar trend of linear correlation was noticed across all 76 libraries.

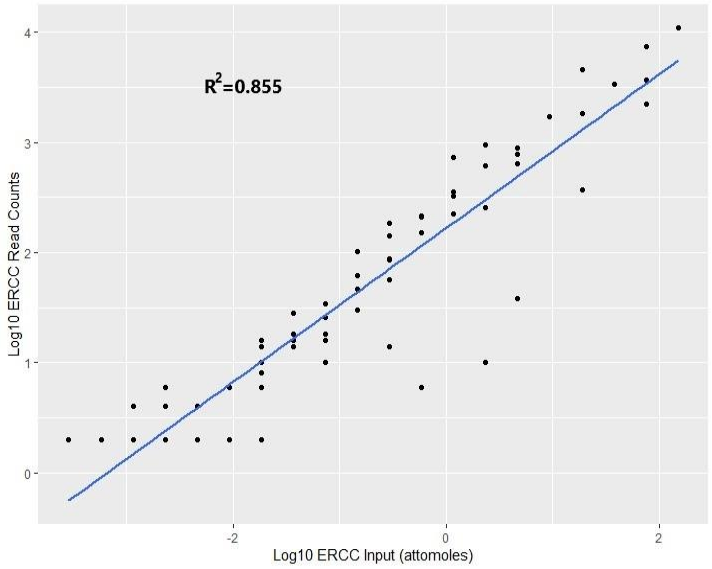


Figure 3.4: Linear correlation plot (R2=0.885) between read counts and input concentration of ERCC spike-in transcripts, determining the library preparation. quality.

The assessment, based on platform fold change response, was done by comparing the observed fold change ratio of Mix1:Mix2 for each ERCC transcript to their expected fold change ratio, and presented a linear correlation between the expected and observed fold change ratios with R^2 value of 0.774 (Figure 3.5). Both RNA correlation

plots revealed no extreme technical outliers, suggesting that the sequencing platform worked and the observed expression differences in the dataset are real.

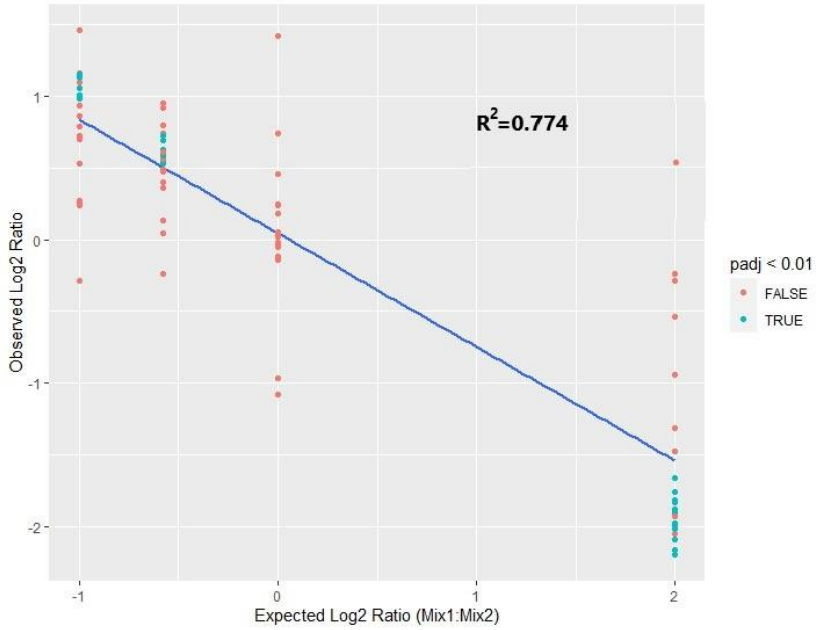


Figure 3.5: Plot showing platform fold change response through a linear relationship between expected and observed fold change ratios of ERCC spike-in transcripts in ERCC Exfold mixes. Data points are shaded based on the significant (padj < 0.01) expression data.

Hierarchical Clustering and Principal Component Analysis

Biological variation among samples, based on the correlation of candidate gene expression, was visualised by PCA. Two types of PCA charts were made: the first was made using read counts of all genes except the ERCC spike-ins and the *Arabidopsis* genes (Figure 3.6) and the second was made using read counts of only differentially expressed genes (p- value < 0.05) between the parental species and reciprocal forms of *T. miscellus* (Figure 3.7). In the first PCA plot (Figure 3.6), PC1 has split the samples by developmental stage and PC2 has separated them by differences in gene expression among species, thus, resulting in clustering of both mid-developmental stage and mid-species (i.e., allopolyploid) in the middle of the sample distribution. The second PCA plot (Figure 3.7) led to clustering of samples within the species as well as a clear splitting of

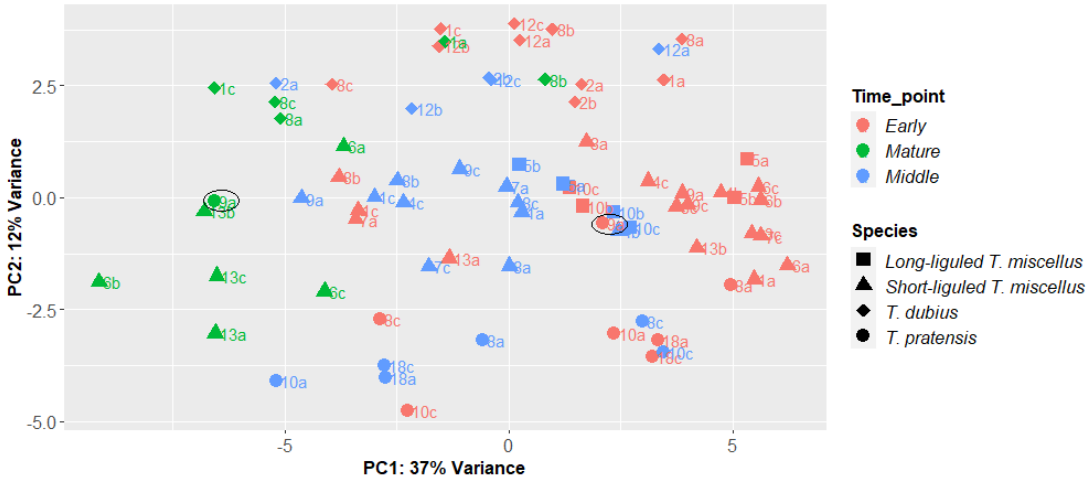


Figure 3.7: PCA plot showing the clustering of individual samples within species at different timepoints (early-, middle- and mature-growth stages). The sample labels with the same digit (1, 5, 8, etc.) but different letter (a, b, c) represent replicates of the same individual. For example, 12a, 12b and 12c represent replicates of one *Tragopogon dubius* individual. The circled samples collected at early- and mature-growth stages, represent the *T. pratensis* individual that showed suspicious behavior and clustered with allopolyploid species.

species, such that the allopolyploid species was between both parent species. However, two samples from the same *T. pratensis* individual, that were collected at the early- and mature-growth stages, demonstrated suspicious behavior and clustered with allopolyploid

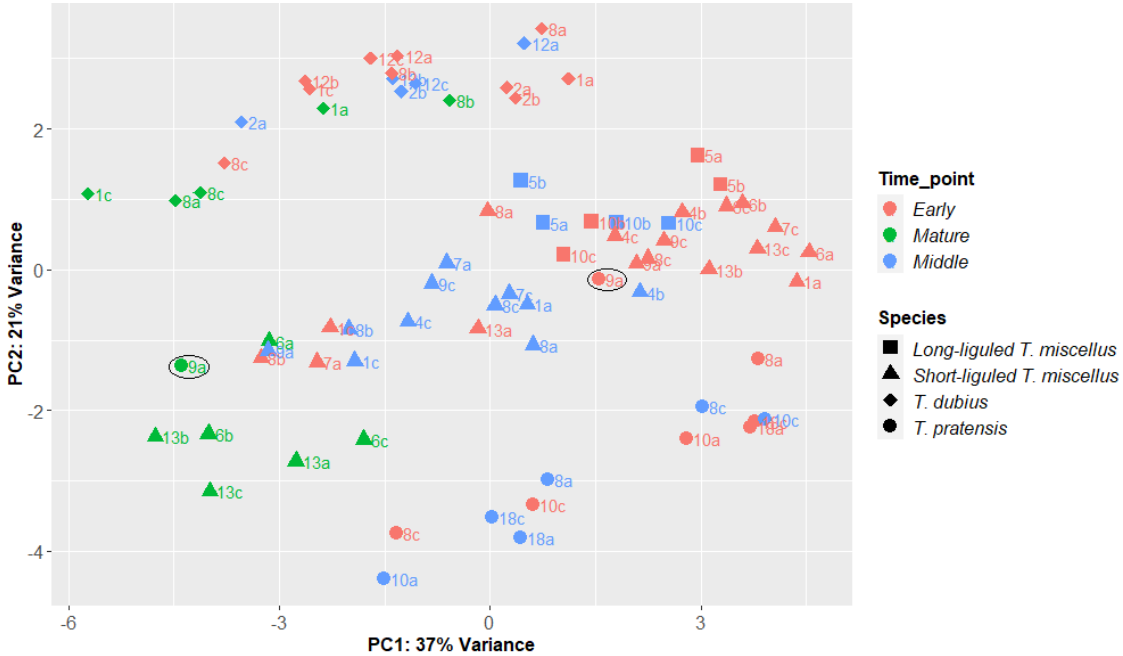


Figure 3.6: PCA of differentially expressed genes ($p < 0.05$) among parents and allopolyploid species at different growth stages. Circled individuals represent *Tragopogon pratensis* samples which clustered with allopolyploid species.

species, encircled in both charts. Because these plants were germinated from field collected seeds, we checked whether this individual could be a hybrid rather than a diploid of *T. pratensis* via PCR of a diagnostic gene (TDF-85) between T. *dubius* and *T. pratensis* (Tate et al. 2006b). The PCR confirmed that this individual was diploid *T. pratensis*, but the placement of this individual is still a mystery (Supplementary Figure 3.4). The sample swap during RNA extractions and library preps may be another plausible explanation. However, RNA from these samples was extracted on different days along with other samples, therefore, we can expect that sample swap during extractions is unlikely. On the other hand, RNASeq libraries for these 2 samples were prepared in the same batch with 24 other samples. If swap or contamination happened during this step, we are not sure at what point this could have occurred as these samples were in separate tubes and extra care was done during all steps. In general, PCA plots displayed clear overall distribution of samples, depicting that the signal in the RNASeq data is biologically accurate with no obvious outliers. This gave confidence that the data are characteristic of the samples and are not due to technical artifacts during library preparation through to the sequencing platform.

Hierarchical clustering analysis was also done to examine the diversity of gene expression profiles and to visualise relationships among genes and samples across all developmental stages in order to identify potential sample outliers. Like PCA, the heatmap generated through hierarchical clustering also confirmed that no apparent outliers existed in the dataset, as shown by overall expression patterns across all samples and developmental stages (Figure 3.8).

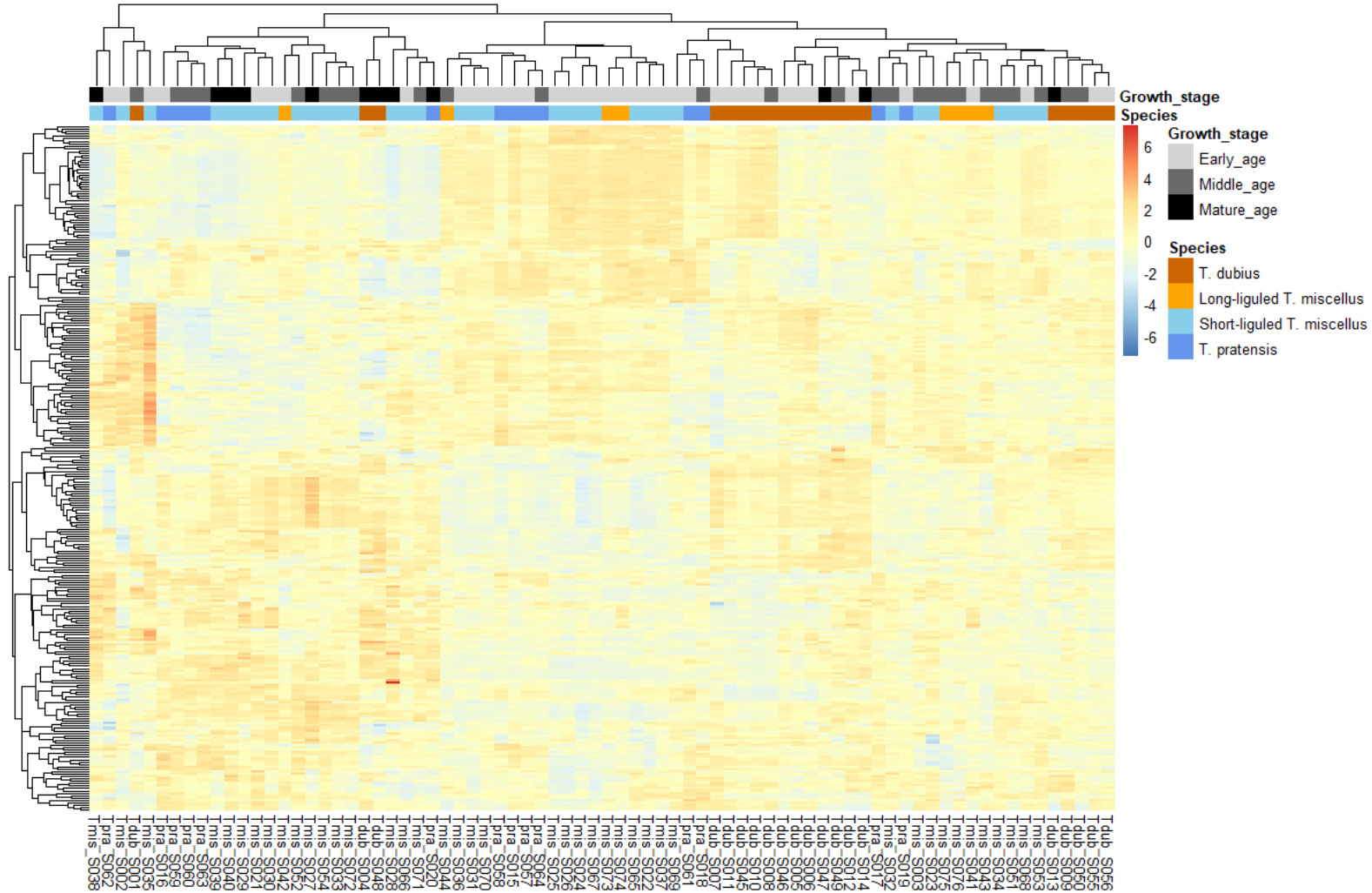


Figure 3.8: Hierarchical clustering of differential expression data of all examined cytonuclear and control genes (not targeted to organelles) of individuals from populations of *Tragopogon dubius*, *T. pratensis* and reciprocal forms of *T. miscellus* at all growth stages. Differential gene expression is displayed in colors ranging from red to blue as shown in the legend. The positive to negative values on the scale bar correspond to up- and down-regulation of gene expression, respectively.

3.4.3. Differential Gene Expression Analysis

Because no obvious outliers were found in the dataset, all samples of respective growth stages from *T. dubius* and *T. pratensis* populations were combined for differential expression analysis. However, all examined populations of both allopolyploid forms were considered independently to identify any differences among their expression patterns, since these populations have formed multiple times in nature. Following differential expression analysis, genes that passed the previously mentioned filtering criteria (section 3.3.5.4) were utilized for downstream analyses. A small number of these genes were found to be significantly differentially expressed between diploid parent species (Supplementary Table 3.2), between each parental species and populations of *T. miscellus* allopolyploids (Supplementary Tables 3.3 & 3.4) and between short- and long-liguled *T. miscellus* populations (Supplementary Table 3.5) as detailed in the following sections.

3.4.4. Total Gene Expression Levels

Differential gene expression analysis between diploid parental species and reciprocal forms of *T. miscellus* was initially done at total gene expression levels, without considering homeolog expression differences. This analysis showed that the great majority of organelle-targeted duplicated nuclear genes (90-98%) demonstrated no expression change in allopolyploids compared to their parent species (Figure 3.9). Those genes (1-3%) which did show differential expression in allopolyploids relative to parent species, exhibited non-additive expression patterns in individuals of late-flowering plants. However, in the case of early-flowering plants, only population 3073 showed such patterns at the early growth stage. These non-additive expression patterns were primarily expression level dominance (Figure 3.9). In long-liguled *T. miscellus*, expression level dominance was unbalanced toward the paternal parent, *T. pratensis*, at both the early- and

middle-growth stages. Likewise, all late-flowering populations of short-liguled *T. miscellus* showed unbalanced expression level dominance toward the paternal parent, *T dubius,* at the early-age growth stage. At the middle-age growth stage, mixed patterns of unbalanced expression level dominance toward either parent (pops. 3049 & 3073) or balanced expression level dominance (pops. 104 & 3053) were observed in the short-liguled populations. In the case of transgressive expression patterns, only short-liguled population 3073 showed transgressive down-regulation at the middle- and early-growth stages of late- and early-flowering plants, respectively. None of the duplicated genes in both reciprocal forms of *T. miscellus* showed transgressive up-regulation. There were certain genes with anomalous gene expression patterns that did not fall into any of the above-mentioned expression categories (Supplementary Figure 3.5 & Supplementary Table 3.6).

The differentially expressed organelle-targeted genes that showed expression level dominance in late-flowering plants encode two chlorophyll A/B binding proteins (photosystem II), light-harvesting complex protein isoform 2.1 (photosystem II), rubredoxin like superfamily protein (OXPHOS complex IV) and Sigma factor E (RNA polymerase PEP) (Supplementary Table 3.7). One of the chlorophyll A/B binding proteins (TragDub19360-RA) showed dominance in short-liguled populations but additive expression in the long-liguled form. In the case of early-flowering plants, two genes (30S ribosomal protein and ribosomal protein 4) from ribosomal genes complexes showed expression level dominance patterns. In the case of transgressive expression patterns, sigma factor E and photosystem II light-harvesting complex gene isoforms (2.2 and 2.3) showed transgressive down-regulation in early- and late-flowering individuals of the short-liguled population 3073, respectively.

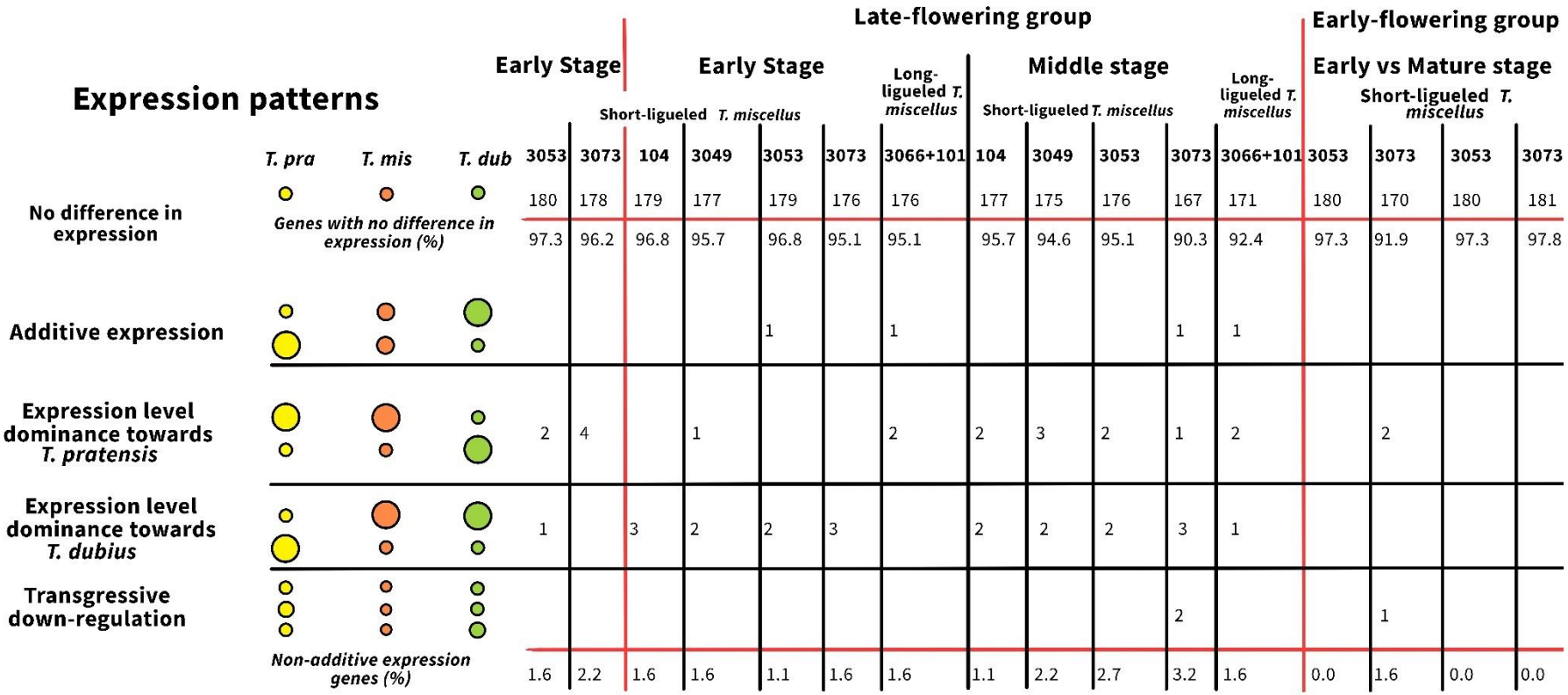


Figure 3.9: Comparison of expression patterns including additive and non-additive expression of 185 organelle-targeted duplicated nuclear genes in reciprocal forms of *Tragopogon miscellus* allopolyploids relative to their diploid parent species. Non-additive expression patterns involve expression level dominance and transgressive expression. *T. pra*, *T. dub* and *T. mis* = *T. pratensis*, *T. dubius* and *T. miscellus*, respectively. The first comparison at the early stage was done considering all early-age samples for both late- and early-flowering plants.

The control nuclear genes and retrograde-signaling partners of dual-targeted genes, which are encoded entirely by the nuclear genome, followed the parental patterns in the case of total gene expression level (Supplementary Figures 3.6 & 3.7).

3.4.5. Homeolog Specific Analyses

Once differential expression analysis at total gene expression levels was done, Homeolog expression differences were determined next in both allopolyploid forms by performing differential expression analysis on reads mapped to homeologs. In contrast to total gene expression levels, a substantial number of homeologs in both reciprocal forms of *T. miscellus* allopolyploid showed differential homeolog expression. These differentially expressed homeologs displayed slightly more bias toward *T. pratensis* across all samples, demonstrating unbalanced homeolog expression bias. In short-liguled *T. miscellus*, 29.7% and 21.8% homeolog pairs showed biased expression toward *T. pratensis* and *T. dubius*, respectively (Table 3.1). In long-liguled *T. miscellus*, 27.3% and 15.3% homeologs displayed bias toward *T. pratensis* and *T. dubius*, respectively (Table 3.1). However, on average, 48.5% and 57.4% of the homeologs in short- and long-liguled *T. miscellus*, respectively, also depicted unbiased expression across all developmental stages (Table 3.1).

The majority of control nuclear homeologs and retrograde-signaling homeologs showed unbiased expression (Supplementary Tables 3.8 & 3.9). The control nuclear homeologs that did show biased expression (40.8%) in short-liguled *T. miscellus* showed mixed patterns of unbalanced homeolog expression bias toward either parent species. However, in the long-liguled form, expression bias toward the paternal parent, *T. pratensis* was observed. Likewise, retrograde-signaling partners showed unbalanced homeolog expression bias toward respective paternal parent in both allopolyploid forms.

Table 3.1: The total number and percentage of organelle-targeted nuclear genes showing unbiased or biased expression toward either parent species (*Tragopogon dubius* or *T. pratensis*) in each population of short- (sm) and long-liguled (lm) *T. miscellus* at early-, middle- and mature-growth stages.

	Short-liguled *T. miscellus*										**Long-liguled *T. miscellus***		**Average % sm/lm**
	Early stage				**Middle stage**				**Mature stage**		**Early stage**	**Middle stage**	
	104	**3049**	**3053**	**3073**	**104**	**3049**	**3053**	**3073**	**3053**	**3073**	**3066+101**	**3066+101**	
Unbiased 	88 (48.6%)	97 (54.2%)	84 (47.2%)	83 (46.6%)	88 (49.2%)	91 (50.8%)	82 (46.3%)	81 (45.8%)	95 (53.1%)	76 (43.2%)	102 (57%)	104 (57.8%)	48.5/57.4
Biased toward *T. pratensis* 	54 (29.8%)	44 (24.6%)	58 (32.6%)	54 (30.3%)	52 (29.1%)	47 (26.3%)	60 (33.9%)	54 (30.5%)	51 (28.5%)	55 (31.3%)	49 (27.4%)	49 (27.2%)	29.7/27.3
Biased towards *T. dubius* 	39 (21.5%)	38 (21.2%)	36 (20.2%)	41 (23%)	39 (21.8%)	41 (22.9%)	35 (19.8%)	42 (23.7%)	33 (18.4%)	45 (25.6%)	28 (15.6%)	27 (15%)	21.8/15.3
Total genes	181	179	178	178	179	179	177	177	179	176	179	180	
Biased gene expression	51.4%	45.8%	52.8%	53.4%	50.8%	49.2%	53.7%	54.2%	46.9%	56.8%	43.0%	42.2%	51.5/42.6

In comparison to expression differences observed between parental genomes, ~44-57% and ~58-60% homeologs displayed parental legacy across all developmental stages in both short- and long-liguled *T. miscellus*, respectively (Figure 3.10). For instance, the most common parental legacy pattern observed in both allopolyploid forms included loci that were not differentially expressed between the parents, followed the same trend in polyploids by displaying unbiased expression. Following polyploidization, only a few genes showed an absence of homeolog expression bias in both allopolyploid forms. In addition, both reciprocal forms established novel homeolog expression bias patterns that were not evident in the parental loci; 41-56% homeologs in short-liguled *T. miscellus* populations versus 38-41% in the long-liguled form (Figure 3.10).

The nuclear-encoded control and retrograde-signaling homeologs that were not targeted to organelles generally showed similar patterns to the cytonuclear genes, such as inherited parental expression patterns and novel homeolog bias patterns following polyploidy (Supplementary Figure 3.8 & 3.9).

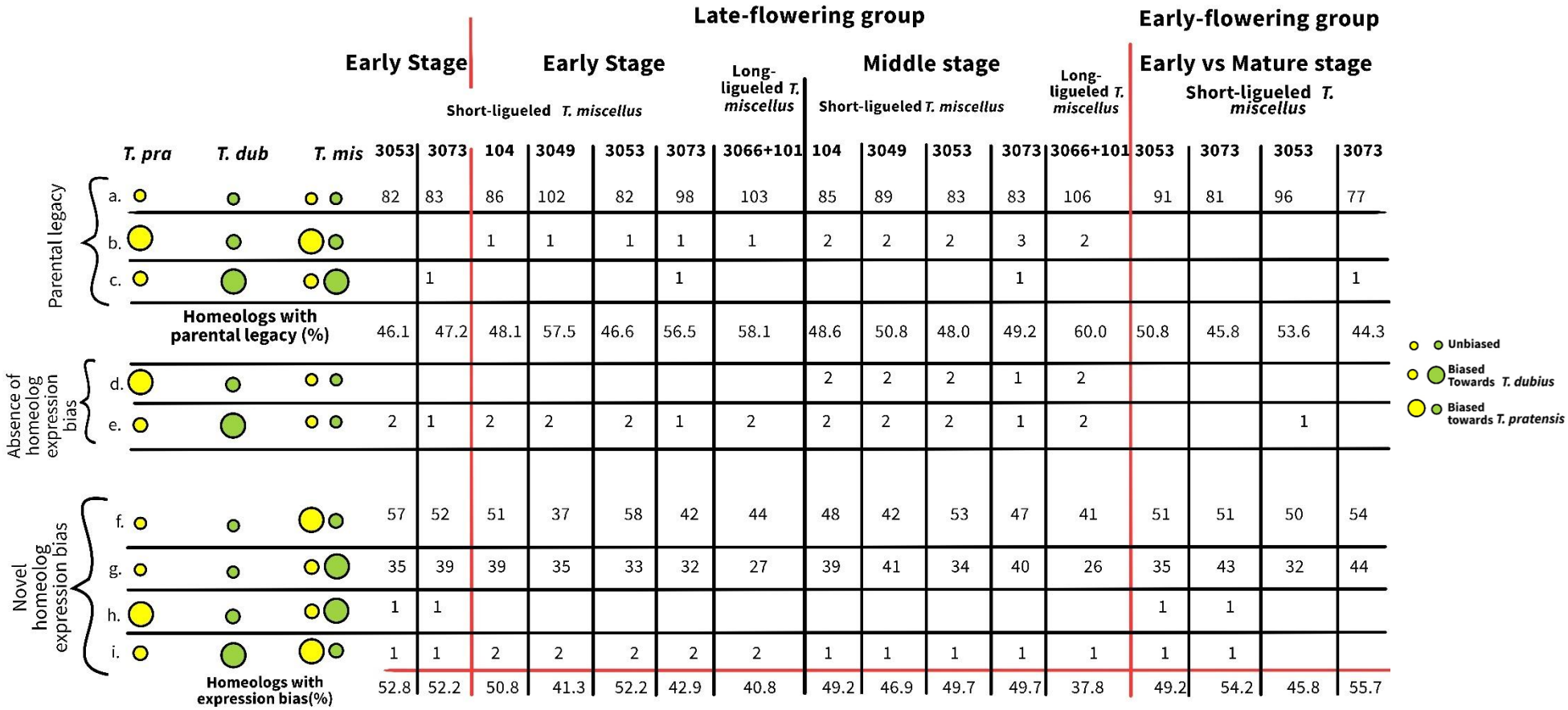

Late-flowering group
Early-flowering group
Early Stage
Early Stage
Long-ligueled T. miscellus
Middle stage
Long-ligueled T. miscellus
Early vs Mature stage
Short-ligueled T. miscellus
Short-ligueled T. miscellus
T. pra
T. dub
T. mis
3053 3073 104 3049 3053 3073 3066+101 104 3049 3053 3073 3066+101 3053 3073 3053 3073
Parental legacy
a. 82 83 86 102 82 98 103 85 89 83 83 106 91 81 96 77
b. 1 1 1 1 1 2 2 2 3 2
c. 1 1 1
Homeologs with parental legacy (%) 46.1 47.2 48.1 57.5 46.6 56.5 58.1 48.6 50.8 48.0 49.2 60.0 50.8 45.8 53.6 44.3
Absence of homeolog expression bias
d. 2 2 2 1 2
e. 2 1 2 2 2 1 2 2 2 2 1 2 1
Novel homeolog expression bias
f. 57 52 51 37 58 42 44 48 42 53 47 41 51 51 50 54
g. 35 39 39 35 33 32 27 39 41 34 40 26 35 43 32 44
h. 1 1 1 1
i. 1 1 2 2 2 2 2 1 1 1 1 1 1 1
Homeologs with expression bias(%) 52.8 52.2 50.8 41.3 52.2 42.9 40.8 49.2 46.9 49.7 49.7 37.8 49.2 54.2 45.8 55.7
Unbiased
Biased Towards T. dubius
Biased towards T. pratensis

Figure 3.10: Comparison of the number of organelle-targeted genes exhibiting homeolog specific expression patterns between polyploids with respect to expression patterns in parent species. Green and yellow circles represent the homeologs derived from *Tragopogon dubius* and *T. pratensis*, respectively. **Parental legacy** (categories a, b & c) is when the expression patterns of genes in the parent species are vertically transmitted in the polyploid species. For example, most of the genes showed no differential expression between parents, thus, the homeologs in polyploids showed unbiased expression patterns (category a). **Absence of homeolog expression bias** means the genes were differentially expressed between parent species but in polyploids this trend was absent and homeologs showed no bias (categories d & e). **Novel homeolog expression bias** represents expression bias towards either parent in polyploids although such patterns were not originally observed in either parent species (categories f, g, h & i). The first comparison at the early stage was done considering all early-age samples for both late- and early-flowering plants.

Homeolog expression patterns of each short-liguled population were also compared to the expression patterns in the long-liguled form. It was noticed that 75-86% and 76-84% of homeologs at early- (Figure 3.11) and middle- (Figure 3.12) growth stages, respectively, displayed similar patterns including unbiased expression or expression bias toward either parent species in each comparison. Nevertheless, only a few homeologs showed directional bias (homeolog expression bias that was altered across reciprocal forms), particularly at the middle stage, i.e., if a certain homeolog was biased toward *T. pratensis* in the short-liguled form, it was shown to be biased toward *T. dubius* in the long-liguled *T. miscellus*. For example, ribosomal protein L31 displayed *T. pratensis* bias in all short-liguled populations and *T. dubius* bias in the long-liguled form (Supplementary Table 3.11). At the early stage, homeologs in short-liguled populations

Figure 3.11: Comparison of the number of organelle-targeted genes showing unbiased and biased homeolog expression toward *Tragopogon pratensis* or *T. dubius* for each population of short-liguled *T. miscellus* and long-liguled *T. miscellus* at the early growth stage. The encircled gene showed directional homeolog bias and the genes in bold red represented the cross-specific maternal bias toward their respective maternal parent.

(9 from 104, 4 from 3049, 16 from 3053 and 15 from 3073) showed bias toward their maternal parent, *T. pratensis* but unbiased expression in the long-liguled form. Likewise, short-liguled populations (8 homeologs from 104, 3 from 3049, 18 from 3053 and 15 from 3073) also demonstrated maternal bias at middle stage. These homeologs that showed maternal bias in all short-liguled *T. miscellus* populations at both early and middle stages but unbiased expression in the long-liguled form include CLP protease 4, ATPase F^0 complex subunit B/B' bacterial/chloroplast and light-harvesting complex of photosystem II 5 (Supplementary Tables 3.10 & 3.11). In the case of long liguled form, relatively small number of homeologs displayed maternal bias toward *T. dubius* but unbiased expression in short-liguled populations at early stage (1 from 104, 5 from 3049, 4 from 3053 and 4 from 3073). For example, chlorophyll A/B binding protein of

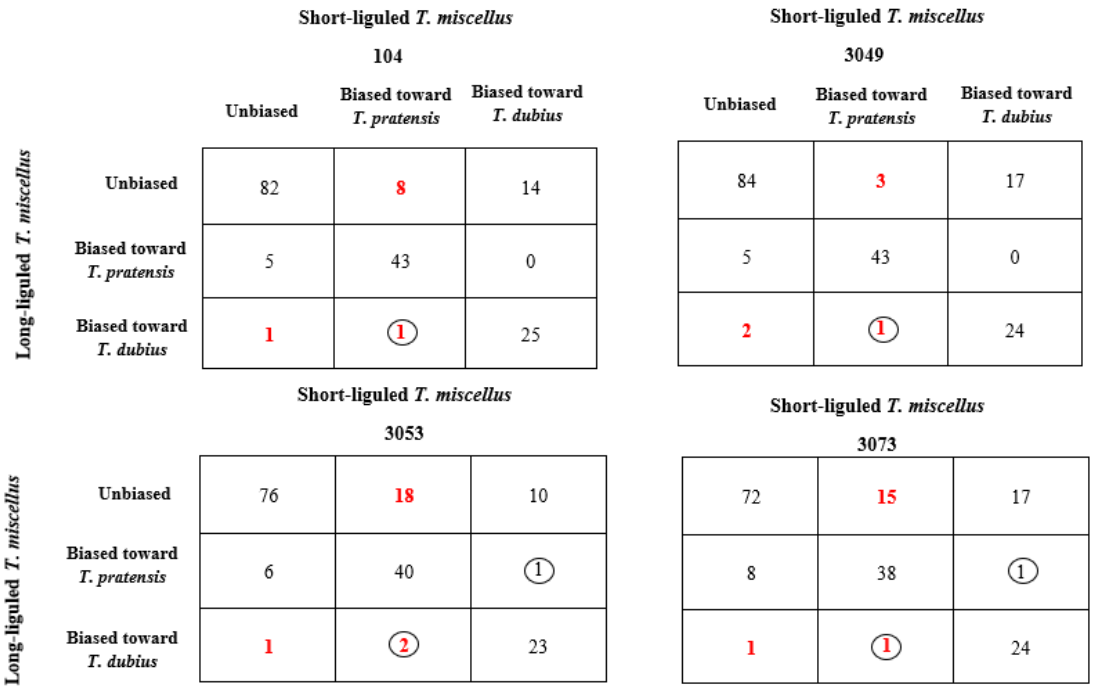


Figure 3.12: Comparison of the number of organelle-targeted genes showing unbiased and biased homeolog expression toward *Tragopogon pratensis* or *T. dubius* between each population of short-liguled *T. miscellus* and long-liguled *T. miscellus* at the middle growth stage. The encircled genes showed directional homeolog bias and the genes in bold red represent the cross-specific maternal bias effect.

photosystem II showed unbiased expression in all short-liguled populations but maternal

bias in the long-liguled form at early stage (Supplementary Table 3.10). Similarly, at the middle stage, homeologs including copper ion binding protein, ubiquinol cytochrome C reductase UQCRX/QCR-9 like family protein and ATP-dependent caseinolytic (Clp) protease/crotonase family protein showed bias toward *T. dubius* in the long-liguled form and unbiased expression in short-liguled populations (Supplementary Table 3.11).

The homeolog specific expression patterns within each population of *T. miscellus* were also examined with respect to developmental stages (Figure 3.13). Most homeologs showed the same expression patterns between growth stages with no directional homeolog bias that was altered across growth stages. Only ~2-6% of the homeologs showed development-specific regulation of homeologs through differential homeolog expression across different growth stages in populations of both short- and long-liguled forms. The list of homeologs that were expressed differentially across growth stages within each population of short- and long-liguled *T. miscellus* is given in the supplementary data (Supplementary Table 3.12). For example, the *mTERF6* gene showed unbiased expression at the early stage, followed by *T. dubius* bias at the middle stage in short-liguled population 3073.

The entirely nuclear-encoded control and retrograde-signaling homeologs generally showed similar expression patterns with a few discrepancies between the reciprocal forms (Supplementary Figures 3.10 & 3.11). Regarding the developmental stages, they generally showed no evidence of growth-specific regulation, except for population 3049 (Supplementary Figures 3.12 & 3.13).

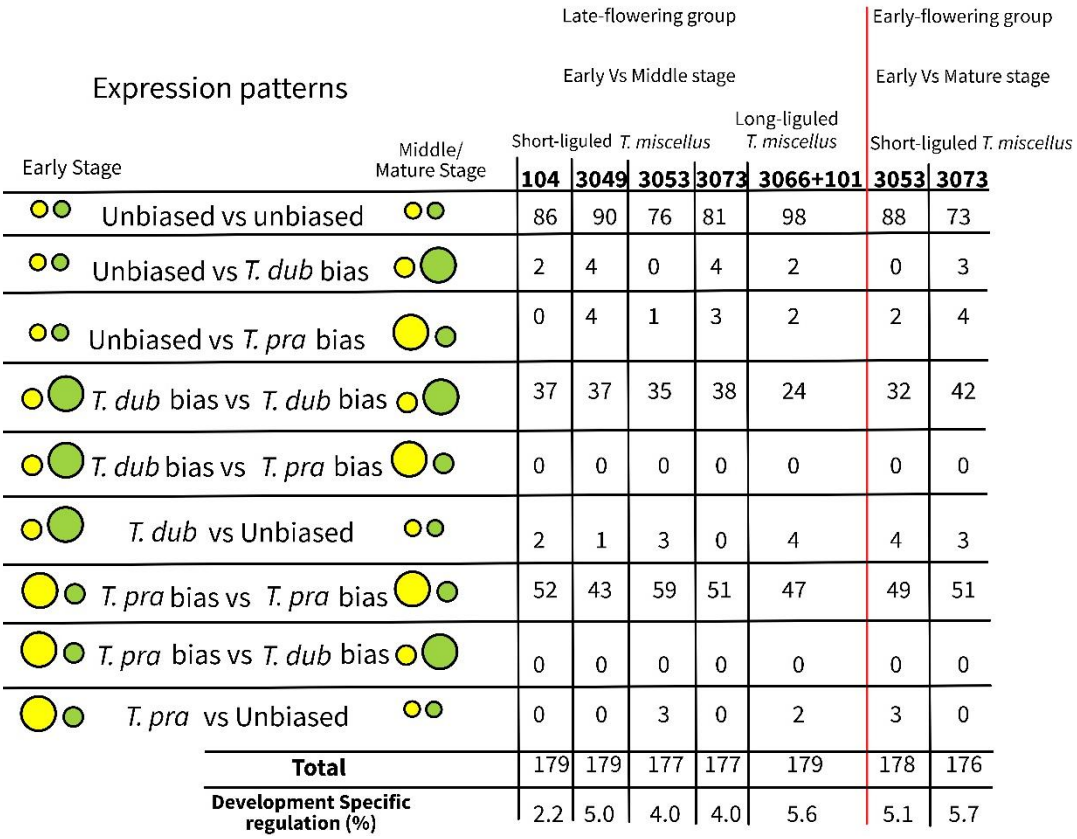


Figure 3.13: Number of organelle-targeted genes showing patterns of unbiased or biased gene expression for early compared to middle- and mature-growth stages in *Tragopogon miscellus* populations. Green and yellow circles represent the homeologs derived from *T. dubius* (*T. dub*) and *T. pratensis* (*T. pra*), respectively. None of the homeologs altered expression bias across developmental stages in both allopolyploid forms.

3.4.6. Cytonuclear Coordination in Rubisco

To confirm how the Rubisco system has coordinated expression in the current study, expression patterns of both small and large subunits of the Rubisco enzyme complex were investigated. Two copies of the small subunit *rbcS, rbcS-A* and *rbcS-B* were retrieved from the *T. dubius* draft genome and eventually used in obtaining transcripts via RNA CaptureSeq. The expression profiles of both copies obtained from RNASeq data for both parent species as well as all individuals of *T. miscellus* were examined. Homeologs of *rbcS-B* copy (TragDub14370) showed unbiased expression in all individuals of short- and long-liguled *T. miscellus*. However, *rbcS-A* copy

(TragDub24397) in all short-liguled populations of the late flowering plant group showed variable homeolog expression patterns. For example, unbiased expression in short-liguled *T. miscellus* population 104, but expression bias toward either *T. dubius* (*T. miscellus* pop. 3049) or *T. pratensis* (*T. miscellus* pops. 3053 & 3073) at both early and middle stages in other short-liguled populations. In the early-flowering group, short-liguled populations, including 3053 and 3073, displayed unbiased expression or *T. pratensis* bias, respectively. However, long-liguled *T. miscellus* individuals exhibited unbiased expression patterns at both early and middle stages for *rbcS-A* copy as well.

To evaluate parental legacy effects on the expression differences in the short-liguled populations, expression patterns of *rbcS* copies were also analysed between the diploid parent species and compared with polyploids. None of the copies were seen to be differentially expressed between parent species at different growth stages.

Moreover, to determine if expression differences of the Rubisco large subunit were regulating or modulating the expression patterns of the small subunits, the expression patterns of organelle-encoded *rbcL* were also determined in the parents as well as the allopolyploid forms. DESeq2 analyses showed that *rbcL* was not differentially expressed between parent species and between reciprocal forms of *T. miscellus*. Additionally, except for short-liguled population 104, all populations of short- and long-liguled forms showed no differential expression relative to their respective maternal parent.

3.4.7. Organelle Gene Expression

Organelle gene expression patterns were also determined in *T. miscellus* allopolyploids to check the role of maternally inherited organelle genes in regulating the cytonuclear coordination through up-regulated expression following polyploidy.

Hierarchical clustering of organelle gene expression showed that the individual samples from populations of both allopolyploid forms clustered with their respective maternal parents (Figure 3.14 and Supplementary Figures 3.14 & 3.15). Additionally, the differential expression analysis of organelle genes between parent species (Supplementary Table 3.13), between short- and long-liguled forms (Supplementary Table 3.14) as well as between allopolyploid populations and their respective maternal parent (Supplementary Table 3.15) indicated that most of the genes have same expression patterns between each comparison. A small number of genes were shown to be differentially expressed in both early and late flowering plant groups.

The ATPase subunit 1 (OXPHOS complex V) was generally differentially expressed in all examined populations of short-liguled *T. miscellus* relative to the long-liguled form (Supplementary Table 3.16). Two genes, nad6 (OXPHOS complex I) and cox2 (OXPHOS complex III) showed up-regulated expression in the long-liguled form with respect to its maternal parent, *T. dubius*. However, none of the genes were up-regulated in short-liguled *T. miscellus* populations with respect to their maternal parent (Supplementary Table 3.16).

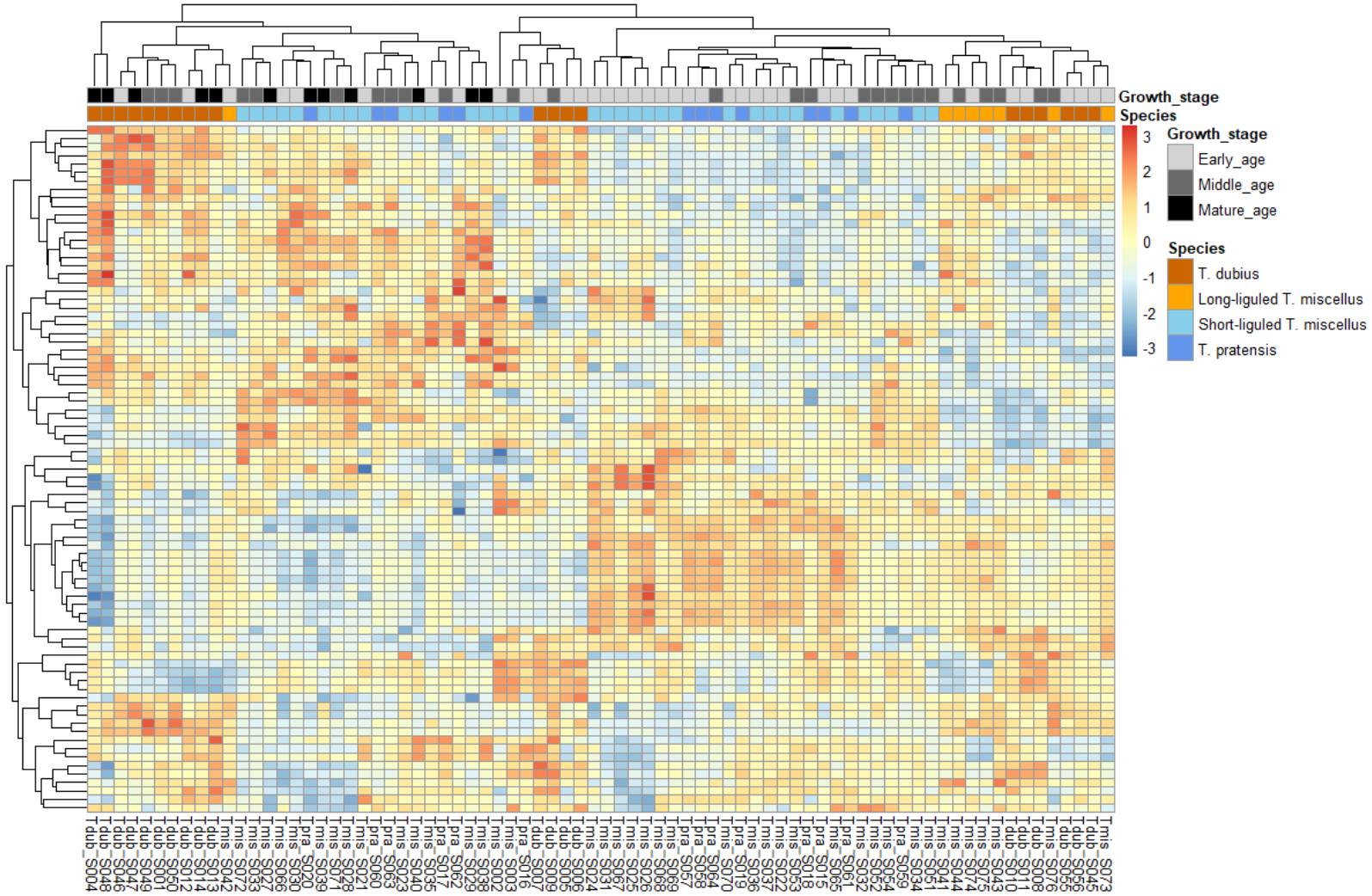


Figure 3.14: Hierarchical clustering of expression data for all organelle genes from individuals of *Tragopogon dubius*, *T. pratensis* and reciprocal forms of *T. miscellus* in both early- and late-flowering plant groups at all growth stages. Differential gene expression is displayed in colors ranging from red to blue as shown in the legend. The positive to negative values on the scale bar correspond to up- and down-regulation of gene expression, respectively.

3.5. Discussion

This study is the first to investigate cytonuclear interactions in natural polyploids of multiple origins (short-liguled) as well as reciprocal forms for multiple interacting cytonuclear complexes. Indeed, very few polyploid systems exist to allow this kind of study, making *Tragopogon* a model system for studying whole genome duplications (Diaz De La Guardia and Blanca 2004; Mavrodiev et al. 2008a; Mavrodiev et al. 2008b; Mavrodiev et al. 2015). *Tragopogon miscellus* is a particularly unique system because the allopolyploids formed multiple times in nature within the last 90-100 years, allowing an early glimpse into allopolyploid genome evolution. Further, the diploid progenitors are known and extant, permitting a direct comparison of the parental expression profiles to those of the polyploids.

Allopolyploidy is an important mechanism in plant evolution that generates genetic diversity and leads to speciation events. Following allopolyploidization, the unequal combination of biparentally inherited nuclear and maternally derived organelle genomes within plant allopolyploids can result in cytonuclear imbalances with subsequent effects on a plant's critical metabolic activities, such as photosynthesis and oxidative phosphorylation (Postel and Touzet 2020). Duplicated nuclear genes undergo several mechanisms to remedy cytonuclear incompatibilities in allopolyploids. These mechanisms may include the immediate return of organelle-targeted genes to single copy (De Smet et al. 2013b; Li et al. 2016), biased loss or rapid evolution of the paternal homeolog (Rand et al. 2004; Sloan et al. 2014; Bock et al. 2014), or biased maternal homeolog expression (Gong et al. 2012, 2014; Sehrish et al. 2015) as was observed in the Rubisco complex of cotton or to some extent in *Tragopogon* allopolyploids. However, contrasting results including paternal bias or a mix of both maternal and paternal bias have also been reported in the Rubisco complex of other allopolyploid systems including

rice, *Cucumis,* and *Brassica* (Wang et al. 2017a; Ferreira de Carvalho et al. 2019; Zhai et al. 2019b; Sharbrough et al. 2022). A study investigating the global gene expression level of several cytonuclear genes including Rubisco, has found varied patterns including paternal and maternal bias in various allopolyploid systems (Grover et al. 2022).

In light of these diverse observations from previous studies that have generally focused on the Rubisco complex, in this study, the expression patterns of duplicated nuclear genes involved in various cytonuclear enzyme complexes (mito-nuclear: five OXPHOS complexes & plastid-nuclear: nine complexes) and dual-targeted genes (three genes) as well as their organelle counterparts were examined via RNA CaptureSeq in *T. miscellus* allopolyploids and representative diploid parents. Moreover, control genes (OXPHOS alternative pathway: eight genes, retrograde-signaling partners: seven genes & housekeeping genes: six genes) that are not targeted to organelle genomes were also included in this expression study to determine how they behave without a cytonuclear context in an allopolyploid system; that is, if such genes would show expression patterns that are different from those of organelle-targeted genes. RNA CaptureSeq was preferred over traditional whole transcriptome RNA sequencing (RNAseq) because this method focuses on the expression of selected transcripts with increased sensitivity and reduced cost or scale, thus permitting analysis of multiple samples. Also, it can characterize low or transiently expressed transcripts that may otherwise be missed using RNAseq. This is the first study to our knowledge that has successfully employed targeted RNASeq approach to investigate gene expression in multiple individuals (76) of allopolyploids and parents.

3.5.1. Cytonuclear Genes May Require Evolutionary Time to Establish Non-Additive Expression Patterns in *T. miscellus* Allopolyploids

To harmonize potential cytonuclear incompatibilities, allopolyploids undergo changes at the total gene expression level (sum of expression of both homeologs) for duplicated nuclear genes (Madlung and Wendel 2013; Yoo et al. 2014a; Grover et al. 2012b; Bao et al. 2019; Doyle et al. 2008; Chen 2007a). In the current study, the total expression levels of duplicated nuclear genes involved in cytonuclear interactions in allopolyploids relative to their diploid parent species were analysed and compared with the expression patterns of control and retrograde-signaling nuclear genes. Our expectation was that organelle-targeted nuclear genes would exhibit non-additive expression patterns/deviation from the parental expression patterns to compensate for cytonuclear conflict. Interestingly, a large of number of organelle-targeted nuclear genes (90-98%) showed no difference in expression patterns between both allopolyploid forms and their parent species. This might be due to little sequence divergence between the diploid parent species and thus between the inherited sub-genomes in the allopolyploids (Sigel et al. 2019). Otherwise, it might be possible that allopolyploids coordinate cytonuclear interactions at the homeolog level instead of at the total gene expression levels as we observed substantial homeolog expression differences in *T. miscellus* allopolyploids (discussed in section 3.5.2-i). Nevertheless, a small percentage of these organelle-targeted genes (1-3%) displayed non-additive expression patterns within *T. miscellus* allopolyploids. In comparison to other older polyploid systems such as cotton (1.5 Mya) and coffee (10,000-50,000 years ago), the low level of non-additive expression patterns has also been observed previously in reciprocal forms of *Tragopogon* and *Polypodium hesperium* (1.2 Mya; Polypodiaceae) allopolyploids, suggesting that non-additive patterns may arise over time, particularly through epigenetic modification (Sigel et al.

2019; reviewed in Yoo et al. 2014b). Moreover, genes duplicated in response to polyploidy have been shown to demonstrate expression divergence at a slower rate compared to tandem duplications (Zou et al. 2009).

Among non-additive expression patterns, expression level dominance is one of the important consequences of allopolyploidy. It occurs when the total expression level of duplicated genes is statistically equal to one of the parent species, and that gene is differentially expressed between parent species. Sub-genome differences in the density of TEs and their methylation status may result in a particular sub-genome being dominant over the other and over evolutionary time, the non-dominant sub-genome could undergo gene fractionation/loss (Hollister and Gaut 2009; Freeling et al. 2012). Because organellar genomes are usually inherited from the maternal parent, expression dominance toward the maternal parent is expected to occur in allopolyploids to avoid cytonuclear incompatibilities (reviewed in Yoo et al. 2014b). This scenario has been analysed in multiple allopolyploid species via whole transcriptome studies with unbalanced expression level dominance toward the maternal or paternal parent generally reported (reviewed in Yoo et al. 2014b). For instance, in *T. mirus* allopolyploids, 7.8% and 8.5% duplicated nuclear genes exhibited expression level dominance toward the paternal (*T. dubius*) and maternal (*T. porrifolius*) parents, respectively (reviewed in Yoo et al. 2014a). Similarly, in *G. hirsutum*, the maternal parent was preferred for expression level dominance (Yoo et al. 2013b). On the other hand, studies in *Coffea* (Bardil et al. 2011) and *Gossypium* (Rapp et al. 2009b) allopolyploids, found dominance toward the paternal parent. Additionally, both reciprocal forms of the fern allotetraploid *Polypodium hesperium* showed unbalanced expression level dominance toward the diploid *P. amorphum* (Sigel et al. 2019), irrespective of the influence of the maternal or paternal

parent. Similarly, the expression level dominance toward *T. pratensis* was observed in both reciprocal forms of *T. miscellus* allopolyploids (Shan et al. 2020a).

So far, only one study to our knowledge has investigated the global gene expression of cytonuclear genes; this study included five unrelated allopolyploid species and found a general bias toward the maternal parent in three of the allopolyploids (*Arachis hypogea*, *Gossypium hirsutum* and *G. barbadense*) (Grover et al. 2022). In the case of duplicated nuclear genes not targeted to organelles, the authors noted the same trend of maternal bias in two of the examined allopolyploids (*Arachis hypogea* and *Gossypium*), while the other systems (*Arabidopsis suecica* and *Chenopodium quinoa*), demonstrated paternal bias. In the current study, duplicated nuclear genes involved in cytonuclear interactions showed unbalanced expression level dominance toward the paternal parent, *T. pratensis*, in long-liguled *T. miscellus* allopolyploids at both early- and middle-growth stages. Similarly, all short-liguled *T. miscellus* populations also exhibited dominance toward their paternal parent, *T. dubius*, at the early-growth stage. However, at the middle-growth stage, short-liguled *T. miscellus* populations exhibited mixed patterns of expression level dominance including balanced (pops. 104 & 3053) as well as unbalanced expression level dominance (pop. 3049) reversed toward the maternal parent. The exception was population 3073, which retained paternal parent dominance from the early stage. Expression level dominance has been shown to be influenced by tissue type in cotton (Yoo et al. 2013b) and environmental factors, such as temperature in *Coffea* (Bardil et al. 2011). For instance, studies in *Coffea* allopolyploids (Bardil et al. 2011; Combes et al. 2013) found that the extent of expression level dominance was changed in response to temperature fluctuations. At 30-26°C, the dominance toward a particular parent (*C. canephora*) was markedly increased compared to low temperature (26-22°C) (Bardil et al. 2011). Likewise, it might be possible that developmental signals are another triggering

factor in switching the direction of expression level dominance in allopolyploids. However, further investigation into the role of developmental signals in the evolution of duplicated genes expression across the same tissue types in reciprocal allopolyploids is needed. Compared to cytonuclear genes, the control nuclear genes, as well as the retrograde-signaling partners of dual-targeted genes, showed similar expression patterns to their parent species in both reciprocal forms of *T. miscellus* allopolyploids.

The expression patterns of organelle-targeted duplicated nuclear genes were also evaluated in terms of transgressive expression and interestingly, no allopolyploid populations showed transgressive expression patterns, except for short-liguled population 3073 (Figure 3.9). Also, there were a few genes that showed anomalous expression patterns that didn't correspond to the additive and non-additive expression patterns in the allopolyploids. These genes generally belong to various cytonuclear complexes including photosystem II, RNA polymerase PEP, Ribosomal genes, and OXPHOS complexes III and IV (Supplementary Figure 3.5 & Supplementary Table 3.6)

Because these *T. miscellus* allopolyploid lineages are evolutionarily young (90-100 years old), they may be in the process of stabilizing their genomes following allopolyploidization and still need more time to establish consistent expression patterns in a particular direction or the degree/magnitude to accommodate cytonuclear interactions. Additionally, the existence of divergent and anomalous expression patterns within *T. miscellus* populations may result from their independent formation, which occurred several times in nature. This could also suggest the evolutionary flexibility following polyploid formation and also that *T. miscellus* allopolyploids may require some time to programme their transcriptome dynamics. On the other hand, there might be other mechanisms operating at homeolog levels to harmonize cytonuclear interactions in allopolyploids, such as sub-genome differences in the densities of transposable elements,

cis- and trans-regulators, small RNAs as well as differential methylation (Wang et al. 2006c; reviewed in Wendel et al. 2018a; Shi et al. 2012b; Hu and Wendel 2019b).

3.5.2. Homeolog Specific Expression Analyses

Homeolog expression bias is another key outcome of whole genome duplication events, which involves the relative contribution of diverged parental homeologs in the allopolyploid total gene expression (Wendel 2000; Jackson and Chen 2010). Like the compensatory mechanisms that might occur to regulate the total gene duplicated expression, the expectation for expression of cytonuclear genes at the homeolog level is that maternal homeologs would be preferred (Gong et al. 2012, 2014; Sharbrough et al. 2017). Likewise, paternal homeologs would be downregulated, lost, or undergo gene conversion events to avoid the likely disagreement between the paternal sub-genome and maternally inherited cytoplasmic genomes (Sharbrough et al. 2022). Such preference for maternal sub-genome/homeologs has been observed in some allopolyploid systems (Gong et al. 2012, 2014), but not all (Wang et al. 2017a; Zhai et al. 2019a; Ferreira de Carvalho et al. 2019), indicating flexibility in the expression profiles of allopolyploids. To assess these expectations, in this study, the relative expression of homeologs in both reciprocal forms of *T. miscellus* allopolyploids were investigated and compared with the expression of their corresponding loci in diploid parent species. The major inferences drawn from this investigation are described below.

i) Maternal bias is not always the rule

Although we expected to find a high percentage of maternal expression bias among the nuclear-encoded homeologs, the majority of the homeologs, 48.5% in short-liguled and 57.4% in long-liguled *T. miscellus* showed unbiased expression across all

growth stages. The remaining homeologs generally displayed unbalanced expression bias toward *T. pratensis* in both short- (51.5%) and long-liguled (42.6%) *T. miscellus* at all growth stages, regardless of the maternal origin. Previous whole transcriptome studies in *T. miscellus* allopolyploids by Buggs et al. (2010c) and Boatwright et al. (2018b) also found that 69% and 48% loci, respectively exhibited no difference between homeolog expression in leaf tissues. Another recent whole transcriptome study conducted using inflorescences of both allopolyploid forms of *T. miscellus* showed that ~80% loci displayed unbiased homeolog expression and loci with biased homeolog expression generally displayed homeolog expression bias toward *T. pratensis* in both allopolyploid forms (Shan et al. 2020a). Similarly, other studies in *Tragopogon* also reported expression bias patterns, with biased expression and genomic loss of *T. dubius* derived homeologs (Tate et al. 2006b, 2009; Buggs et al. 2010c).

On the other hand, similar studies focused on cytonuclear genes by Grover et al. (2022) and Wang et al. (2017a) stated mixed patterns of expression bias toward either the paternal or maternal parent in *Arachis hypogea*, *Gossypium*, *Arabidopsis suecica*, *Chenopodium quinoa* and rice allopolyploids. Furthermore, Grover et al. (2022) noticed that for all investigated homeolog pairs, allopolyploids showed paternal bias. However, some cytonuclear gene categories showed maternal bias in a few individuals (Grover et al. 2022).

Compared to the cytonuclear genes, the control nuclear genes typically showed unbiased expression and a few homeolog pairs with biased expression displayed unbalanced expression bias toward either parent species or balanced homeolog expression bias (pop. 104) at all growth stages in the short-liguled form. In long-liguled *T. miscellus*, homeolog pairs showed bias toward *T. pratensis*, the paternal parent. The direction of bias toward a particular parent species was maintained at each growth stage. However,

the number of loci increased for that bias at later growth stages in short-liguled *T. miscellus* populations. The retrograde-signaling partners of dual-targeted genes depicted paternal bias in both forms.

The occurrence of such stochastic/diverged homeolog expression patterns in different gene categories may suggest that *T. miscellus* allopolyploids achieve cytonuclear coordination through differential regulation of homeologs but keeping the total gene expression levels same as those of diploid parent species. The differential homeolog regulation may involve gaining novel expression bias patterns following polyploidy as discussed in the following section. The homeolog specific regulation could have arisen from the inherited cis- and trans-regulatory differences between parental sub-genomes within an allopolyploid system (Wang et al. 2006c; Shi et al. 2012b; Hu and Wendel 2019b). Moreover, the varying distribution and silencing of transposable elements between parental sub-genomes may play role in establishing such diverse biased homeolog expression patterns in allopolyploids (Wendel et al. 2018b). The impact of TEs in down-regulating homeologous expression and preferential homeologous loss has been observed in various allopolyploids, which eventually leads to biased fractionation over evolutionary time (reviewed in Wendel et al. 2018b). Moreover, the genetic background of independently formed *T. miscellus* populations may influence the expression of genes due to the varying genetic contribution of diploid parent species to these different populations (Symonds et al. 2010). Also, other factors such as the divergence time of the diploid parent species and the age of the allopolyploid may influence the evolution of homeolog expression. Because *T. miscellus* allopolyploids are young, they may have not yet fixed consistent expression patterns following polyploidization, thus the majority of parental homeologs are expressed equally. Unbalanced bias toward *T. pratensis* could

emerge from the aspect that this sub-genome might have more advantageous features, such as a lower TE density and dynamic epigenetic attributes compared to *T. dubius*.

ii) Novel expression patterns exist following polyploidy, even though young *T. miscellus* allopolyploids maintained their parents' expression marks

The availability of extant diploid parent species, *T. pratensis* and *T. dubius*, offers an opportunity to directly compare the effect of parental gene expression on the relative homeolog expression profiles in reciprocal forms of *T. miscellus* allopolyploids, in contrast to other systems where one of the parents is extinct. This situation allowed us to determine if allopolyploids inherited expression differences from their parents. To confirm this, the homeolog expression patterns in reciprocal forms of *T. miscellus* were compared with the expression patterns in both diploid parent species. Approximately, 44-57% and 58-60% homeolog pairs primarily followed their parents' footsteps in short- and long-liguled forms, respectively. Similarly, a majority of loci showed parental legacy in inflorescence tissues of short- (68.7%) and long-liguled (69.3%) *T. miscellus* (Shan et al. 2020a). Other studies in natural *Gossypium* and *Brachypodium* allopolyploids also reported that ~63% and ~65% of examined loci exhibited parental legacy, respectively (Yoo et al. 2013b; Takahagi et al. 2018a). The percentage of allopolyploid loci inheriting parental expression patterns in the current study is comparable to the above-mentioned studies, because this study focused on 293 cytonuclear genes instead of whole RNASeq as performed in other studies.

The remaining loci in this study were shown to diverge from parental expression patterns following allopolyploidization. For instance, novel expression patterns were found in 41-56% and 38-41% of loci in both short- and long-liguled forms, respectively.

The prominent novel expression patterns showed by the genes in *T. miscellus* allopolyploids, included homeolog expression bias toward either parent species while the corresponding loci in parent species demonstrated equal expression levels. Interestingly, most populations in both the early- and late-flowering plant groups showed decline in the percentage of these novel expression patterns in later growth stages, except for 3073 and 3049 populations in the late-flowering group. Likewise, Shan et al. (2020a) reported 16% of loci displayed novel homeolog bias patterns in both reciprocal forms of *Tragopogon* allopolyploids. The control genes as well as retrograde-signaling genes also displayed similar patterns including parental legacy and novel homeolog expression bias like the cytonuclear genes. This evolution of novel homeolog specific expression patterns in *T. miscellus* allopolyploids suggests that polyploidy may have started rewiring the homeolog expression patterns in young allopolyploids to achieve stability through rapidly utilizing the newly generated genetic diversity.

iii) Developmental specific functional diversification of homeologs is yet to be established in young allopolyploids

Developmental or tissue-specific regulation of homeologs have been reported in several allopolyploids (Wang et al. 2006d; Adams et al. 2003a; Chaudhary et al. 2009; Buggs et al. 2011b), but these studies did not focus on cytonuclear complexes. These studies found that in allopolyploids one homeolog was expressed at a certain developmental stage or in a particular organ, but at a later growth stage the alternate homeolog was expressed. These homeologs could either retain their ancestral functions or one of them could develop a novel function (neofunctionalization) (Chaudhary et al. 2009). They may also undergo partitioning in their ancestral function, leading to sub-functionalization (Chaudhary et al. 2009). To balance cytonuclear interactions, it might

be expected that cytonuclear homeologous pairs may also undergo development mediated regulation, such that both homeologs are differentially expressed at distinct growth stages. In light of this hypothesis, the expression patterns of homeolog pairs were compared between early- and middle-growth stages as well as early- and mature-growth stages within each population of both early- and late-flowering plants. Our results showed that the majority of the homeologs of cytonuclear complexes, as well as the control and retrograde genes, in both reciprocal forms predominantly maintained the same expression patterns between developmental stages. Only ~2-6% of the cytonuclear genes showed growth-specific differential homeolog regulation in the young *T. miscellus* allopolyploids. This suggests that functional diversification of homeologs might not have established immediately in response to whole genome duplication events in these young allopolyploids and instead may take several generations to arise. Additionally, as discussed in the earlier section there might be other regulatory mechanisms that need to be investigated in allopolyploids with respect to cytonuclear interactions.

iv) Duplicated nuclear genes implicated in cytonuclear interactions may evolve differently in reciprocally formed *T. miscellus* allopolyploids

T. miscellus allopolyploids offer a rare opportunity to investigate the likely differential regulation of duplicated nuclear genes implicated in cytonuclear interactions because of reciprocally inherited maternal cytoplasmic genomes. Therefore, in this study, the homeolog specific expression patterns among populations of short- and long-liguled forms were also studied. Mostly genes were observed to display similar expression patterns between both reciprocal forms across all growth stages. Although homeolog expression was generally biased toward *T. pratensis* in both allopolyploid forms, a reasonable proportion of homeologs displayed expression bias differences including

maternal bias toward their respective maternal parents, directional bias (homeolog bias was altered between reciprocal forms), etc. between both reciprocal forms. Comparable results regarding variation at the homeolog expression level have also been reported for reciprocal forms of *Polypodium hesperium* (Sigel et al. 2019). Although the expression differences between reciprocal forms of *T. miscellus* allopolyploids are not extensive, this suggests that certain lineage-specific mechanisms or modifications may be working to impose these expression differences in certain gene groups. These could include the evolutionary differences between cytoplasmic genomes or even between the nuclear sub-genomes as discussed above in this section.

3.5.3. Cytonuclear Coordination in Rubisco

Rubisco is a cytonuclear enzyme complex that participates in the carbon assimilation process in plants. It is composed of two subunits, small subunit *rbcS* and large subunit *rbcL,* which are encoded by nuclear and plastid genomes, respectively. The small submit, *rbcS* has been investigated in multiple allopolyploid systems with respect to cytonuclear coordination (discussed in section 3.2) and diverse patterns of expression including maternally biased expression (Gong et al. 2012, 2014), as well as a combination of maternally and paternally biased expression (Wang et al. 2017a; Zhai et al. 2019a) have been observed. In this study, mixed patterns of expression bias toward either parent species, as well as unbiased expression have been observed. The small number of individuals showing maternal bias is consistent with previous results by Sehrish et al. (2015) that cytonuclear coordination in *Tragopogon* allopolyploids may be a slower process and does not occur immediately following polyploidy. Otherwise, it may indicate that the parent species are not too divergent in their genomes and therefore the potential cytonuclear incompatibilities in allopolyploids are minimal for certain cytonuclear

complexes, including this enzyme system, but may differ for others. For example, the divergence of *rbcS*-1 sequences was low between the diploid parent species of *T. miscellus* allopolyploids (Sehrish et al. 2015) compared to other genera including *Arabidopsis*, *Gossypium* and *Triticum* (Gong et al. 2012; Krebbers et al. 1988; Sasanuma 2001). Also, plastid *rbcL* is highly conserved in *Tragopogon*, for instance, *rbcL* in both *T. dubius* and *T. pratensis* species differed by only one SNP that resulted in a synonymous substitution (Sehrish et al. 2015). This indicates that the existence of differential homeolog expression in a few short-liguled populations might occur at the post-transcriptional or post-translational level. For instance, in *Cucumis* allopolyploids, it has been predicted that the paternal *rbcS* homeolog has a higher binding affinity to *rbcL,* which ultimately resulted in strong interactions between the respective proteins, explaining the existence of expression bias toward paternal *rbcS* homeolog in *Cucumis* allopolyploids (Zhai et al. 2019a). Regulation at the translational level or post-transcriptional/translational levels could also be one of the reasons behind the lack of developmental mediated regulation of *rbcS* homeologs, because homeologs have maintained similar patterns across developmental stages, suggesting that developmental regulation is not yet playing a role in cytonuclear coordination of Rubisco complex.

3.5.4. Compensatory Response of Organelle Genes to Polyploidy

Organelle genomes may function as a key regulator in dictating the evolution of nuclear genomes as well as their expression via contributing to the compensatory evolution in the nuclear genome (Osada and Akashi 2012a; van der Sluis et al. 2015). Therefore, any evolutionary differences present between the organelle genomes of parent species might be vertically transmitted into the reciprocal forms of allopolyploids. This in turn could impact the expression of their interacting duplicated nuclear genomes in

allopolyploids. For instance, it is expected that the maternally inherited organelle genome might favor the expression of the maternally inherited nuclear genome. On the other hand, selection may put pressure on the paternal homeologs for deleterious mutations to avoid any potential conflict between them and organelle genomes. Alternatively, organelle genomes may respond by elevating their copy number per organelle or via more organelles per cell (Gyorfy et al. 2021; He et al. 2021; Coate et al. 2020; Murti et al. 2012), which subsequently may result in modulation of organelle expression to coordinate cytonuclear interactions (Doyle and Coate 2019; Coate et al. 2020).

In this study, the differential expression analysis between parental organelle genomes showed that the majority of the organelle genes showed similar expression patterns between both parent species. Likewise, organelle genes in the allopolyploids generally showed no departure from their maternal expression patterns and also exhibited similar patterns between both reciprocal forms. Only a small number of genes displayed differential expression among polyploid populations and in comparison, to parent species. Such slight variations might suggest that organelle genomes in *Tragopogon* are conserved, for instance, the plastid genomes of both *T. dubius* and *T. pratensis* are 99.5% similar (unpublished data). Thus, they have not yet expressed extreme alterations in their gene expression and ultimately did not induce changes in their corresponding nuclear genomes in allopolyploids.

Chapter 4

Investigation of Morphological Trait Variation in Reciprocally Formed *T. miscellus* Allopolyploids

4.1. Abstract

Polyploidy has the capability to alter phenotypic traits due to the newly generated genetic diversity and transcriptomic plasticity in polyploids. The repeated formation of morphologically distinct reciprocal *T. miscellus* allopolyploids provides an excellent system to investigate the phenotypic consequences of allopolyploidy. Extensive genomic, genetic and transcriptomic consequences of allopolyploidy have been explored previously in the *T. miscellus* allopolyploids. In the current pilot study, the effect of allopolyploidy on the morphological traits of *T. miscellus* allopolyploids were investigated. Our results showed that both reciprocal forms of *T. miscellus* generally showed faster germination rates and higher germination percentages relative to their diploid parent species, *T. dubius* and *T. pratensis*. In terms of leaf morphological parameters including leaf number and length, *T. miscellus* allopolyploids had lower leaf density but longer leaves as compared to both parental species. These preliminary findings suggest the potential role of allopolyploidy on the evolution of phenotypic traits of *T. miscellus* allopolyploids. Thus, *T. miscellus* allopolyploids are valuable targets to further explore the evolution of trait plasticity in polyploids.

4.2. Introduction

Whole genome duplication or polyploidy has played an important role in shaping plant diversity and the formation of new species (Adams and Wendel 2005). It has frequently occurred in plants and data suggest that 30-80% angiosperms are polyploids, including economically important crops such as wheat, cotton, coffee, sugarcane and potato (Grant 1981a; Wendel and Doyle 2005; Tate et al. 2005b; Masterson 1994). The high prevalence of polyploidy in plants suggests their evolutionary and ecological success due to the presence of increased genetic material. The greater genetic diversity in polyploids leads to genomic rearrangements, epigenetic remodelling and changes in gene expression in polyploids (Comai 2005a). These genetic and transcriptomic alterations in turn have impact on the morphological, anatomical and physiological characteristics of plant polyploids (Levin 2004; Chen 2007b; Gaeta et al. 2007). For example, plant allopolyploids have been noted to have superior traits compared to their parental species, including increased growth (fruit, leaf and seed size), altered photosynthetic and transpiration rates, stress tolerance, disease resistance and ecological diversification (Miller et al. 2012; Zhao et al. 2005; Wu et al. 2012; Zhang et al. 2015a; Ha et al. 2009; Wendel and Cronn 2003b; Otto and Whitton 2000; Maherali et al. 2009). The better growth of allopolyploids relative to their diploid parent species is the outcome of hybrid vigour or heterosis, resulting from the interspecific hybridization between parental species in allopolyploids (Chen 2010). The hybrid vigour eventually becomes fixed following a whole genome duplication event in allopolyploids, leading to advantageous traits and higher fitness than diploid species (Chen 2010). Because of these beneficial traits, plant polyploids are of great interest for plant breeders. Therefore, understanding the phenotypic consequences of polyploidy is of paramount importance in evolutionary biology.

The classic morphological response to polyploidy is increase in the size of cells due to the greater number of duplicated chromosomes, which consequently may result in the larger organs or tissues (Beaulieu et al. 2008; Otto and Whitton 2000). For example, polyploids may have larger and thicker leaves, stomatal guard cells, flowers, pollen grains and fruits compared to their diploid parent species (Sugiyama 2005; Chen et al. 2009; Balao et al. 2011; Bomblies 2020; Snodgrass et al. 2017). Also, polyploidy can positively impact seed structure, resulting in larger and heavier seeds with faster germination rates and more viability (higher germination percentages) (Hoya et al. 2007; Astuti et al. 2020; Stevens et al. 2020). Besides morphological changes, genome plasticity could also result in physiological variations in polyploids including improved stress responses to biotic and abiotic stresses and changes in photosynthetic rates (Sattler et al. 2016; Arias and Bhatia 2015; Sun et al. 2015). For instance, increased drought tolerance was found in *Lonicera japonica* and *Citrus* tetraploids compared to diploids (Ruiz et al. 2016; Li et al. 2009). The altered responses to abiotic stresses between polyploids and diploids are associated with changes in the leaf anatomy, including fewer but larger stomata, and thicker upper and lower epidermis (Li et al. 2009). Because polyploids have larger leaves and stomata as well as a higher content of photosynthetic pigments, it is apparent that they could have higher photosynthetic rates. However, in comparison to diploid parent species, the photosynthetic rates in polyploids are species dependent. For example, elevated photosynthetic rates were observed in synthetic autotetraploids of *Malus* × *domestica* (Xue et al. 2017) and *Phlox drummondii* (Vyas et al. 2007) compared to their diploid parent species. The increased photosynthetic capacity might be the outcome of increased concentration of photosystems, resulting from high chlorophyll a/b ratios in polyploids, as reported by Austin et al. (1982). On the contrary, *Fragaria* tetraploids and *Triticum* tetra- and hexaploids showed lower photosynthetic rates relative to diploid

species (Hejnák et al. 2016; Gao et al. 2017). The same study by Hejnák et al. (2016) noticed that *Triticum* hexaploids, except *Triticum aestivum,* had even lower photosynthetic rates than tetraploids. Likewise, several other studies on different ploidies of *Triticum* and citrus trees have shown the negative correlation of ploidy levels with the net photosynthetic rates (Warner and Edwards 1993; Bilgrami et al. 2015; Romero-Aranda et al. 1997; Srivalli and Khanna-Chopra 2004; Maosong et al. 2008). This might have resulted from the decreased stomatal conductance and transpiration rates due to the lower density of stomata in polyploids, which ultimately slowed down the photosynthetic process and the absorption of intracellular CO2 (Maosong et al. 2008). Moreover, the large size of mesophyll cells in polyploid leaves provides less surface area to volume ratio, introducing an increased barrier to CO2 exchange (Srivalli and Khanna-Chopra 2004). However, the examined polyploids (4*x*, 5*x* and 6*x*) of *Allium oleraceum* showed no significant difference in net photosynthetic rates and stomatal density (Ježilová et al. 2015). Lastly, allopolyploids (tetraploid to decaploid) of *Festuca arundinacea* showed an increase in photosynthetic rates with ploidy level (Warner and Edwards 1993). These species-specific differences suggest the lack of a universal trend or behaviour in polyploids.

Tragopogon miscellus is an excellent natural allotetraploid system (*T. dubius* x *T. pratensis*) for studying phenotypic consequences of allopolyploidy because it exists in two morphologically distinct reciprocal forms, short- and long-liguled *T. miscellus*. In long-liguled *T. miscellus*, *T. dubius* is the maternal parent and it has more relaxed inflorescences and long ligules, however, the short-liguled form with *T. pratensis* as the maternal parent has more constricted inflorescences and short ligules (Figure 4.1) (Ownbey 1950; Ownbey and McCollum 1953; Soltis and Soltis 1989). Past studies have demonstrated extensive chromosomal, genomic and transcriptomic variation in

Tragopogon allopolyploids including intergenomic translocations, homeolog loss and silencing, differential homeologous expression, and differential proteomes (Buggs et al. 2012a; Buggs et al. 2011a; Tate et al. 2009a; Tate et al. 2006a; Buggs et al. 2009a; Chester et al. 2012a; Spoelhof et al. 2017).

Figure 4.1: Inflorescence differences between reciprocally formed *Tragopogon miscellus* allopolyploids. Left side shows relaxed inflorescence of long-liguled *T. miscellus* and right side shows constricted inflorescence of short-liguled form.

In terms of the genetic mechanisms causing morphological differences between reciprocal forms of *T. miscellus*, Shan et al. (2020b) investigated the floral transcriptomes of both reciprocal forms and speculated that genes involved in pectin metabolism might be playing a role in the floral patterning and morphological differences between short and long-liguled forms of *T. miscellus*. In this pilot study, we investigated potential differences in the germination rates and compared leaf morphology between reciprocal forms of *T. miscellus* and diploid parent species to test following hypotheses:

i) *T. miscellus* allopolyploids will have faster germination rates and higher germination percentages compared to those of the parent species.

ii) *T. miscellus* allopolyploids will have more and larger leaves relative to their parent species.

4.3. Materials and Methods

4.3.1. Plant Materials

The diploid parental species, *Tragopogon dubius* and *T. pratensis*, as well as short- and long-liguled *T. miscellus* allopolyploids were used to assess the morphological traits. Seed collected from field sites in the USA (2016 collection) were used in the study. Initially, three seeds for each lineage (one individual from a population) were started and in total, the sample set included 156 plants from all species (Table 4.1). A subset of these plants was included in the expression work and leaf tissues were collected after morphological assessments (details in Chapter 3, section 3.3.1 & Supplementary Table 3.1).

4.3.2. Seed Germination Traits

Seed germination traits, including seed germination rates and percentages, were assessed in diploid parental species and reciprocal forms of *T. miscellus* allopolyploids. In terms of germination percentages, the germination success rate of seeds was calculated for each species after 14 days of sowing to estimate differences in the viability of seeds. In addition, seed germination rates of each species were also estimated based on the first sign of cotyledon emergence after one week of sowing and then every second day until 14 days of sowing.

Table 4.1: List of individuals from natural populations of diploid parent species, *Tragopogon dubius* and *T. pratensis* as well as reciprocal forms of *T. miscellus* included in this study.

Species	Population	Population ID	Individual ID
T. dubius	Pullman, Washington, USA	3040	3
T. dubius	Pullman, Washington, USA	3040	10
T. dubius	Pullman, Washington, USA	3040	4
T. dubius	Pullman, Washington, USA	3040	2
T. dubius	Moscow, Idaho, USA	106	1
T. dubius	Moscow, Idaho, USA	106	2
T. dubius	Moscow, Idaho, USA	106	4
T. dubius	Moscow, Idaho, USA	106	5
T. dubius	Spangle, Washington, USA	3050	1
T. dubius	Spangle, Washington, USA	3050	3
T. dubius	Spangle, Washington, USA	3050	4
T. dubius	Spangle, Washington, USA	3050	2
T. dubius	Rosalia, Washington, USA	3072	6
T. dubius	Rosalia, Washington, USA	3072	4
T. dubius	Rosalia, Washington, USA	3072	5
T. dubius	Rosalia, Washington, USA	3072	7
T. dubius	Oakesdale, Washington, USA	3055	8
T. dubius	Oakesdale, Washington, USA	3055	12
T. dubius	Oakesdale, Washington, USA	3055	11
T. dubius	Oakesdale, Washington, USA	3055	15
T. pratensis	Oakesdale, Washington, USA	3052	7
T. pratensis	Oakesdale, Washington, USA	3052	17
T. pratensis	Oakesdale, Washington, USA	3052	18
T. pratensis	Oakesdale, Washington, USA	3052	6
T. pratensis	Garfield, Washington, USA	3058	9
T. pratensis	Garfield, Washington, USA	3058	8
T. pratensis	Garfield, Washington, USA	3058	10
T. pratensis	Garfield, Washington, USA	3058	7
T. miscellus (long-liguled)	Pullman, Washington, USA	3066	11
T. miscellus (long-liguled)	Pullman, Washington, USA	3066	10
T. miscellus (long-liguled)	Pullman, Washington, USA	3066	15
T. miscellus (long-liguled)	Pullman, Washington, USA	3066	14
T. miscellus (long-liguled)	Pullman, Washington, USA	101	4
T. miscellus (long-liguled)	Pullman, Washington, USA	101	5
T. miscellus (long-liguled)	Pullman, Washington, USA	101	6
T. miscellus (long-liguled)	Pullman, Washington, USA	101	10
T. miscellus (short-liguled)	Spangle, Washington, USA	3049	1
T. miscellus (short-liguled)	Spangle, Washington, USA	3049	2
T. miscellus (short-liguled)	Spangle, Washington, USA	3049	3
T. miscellus (short-liguled)	Spangle, Washington, USA	3049	4
T. miscellus (short-liguled)	Garfield, Washington, USA	104	2
T. miscellus (short-liguled)	Garfield, Washington, USA	104	4
T. miscellus (short-liguled)	Garfield, Washington, USA	104	1
T. miscellus (short-liguled)	Garfield, Washington, USA	104	3
T. miscellus (short-liguled)	Rosalia, Washington, USA	3073	4
T. miscellus (short-liguled)	Rosalia, Washington, USA	3073	5
T. miscellus (short-liguled)	Rosalia, Washington, USA	3073	6
T. miscellus (short-liguled)	Rosalia, Washington, USA	3073	7
T. miscellus (short-liguled)	Oakesdale, Washington, USA	3053	2
T. miscellus (short-liguled)	Oakesdale, Washington, USA	3053	8
T. miscellus (short-liguled)	Oakesdale, Washington, USA	3053	9

Species	Population	Population ID	Individual ID
T. miscellus (short-liguled)	Oakesdale, Washington, USA	3053	13

4.3.3. Leaf Morphological Traits

The plants from each lineage of diploid parents and allopolyploids were observed after 9 weeks of germination for leaf morphological parameters. *T. dubius* populations: 3040 and 3072, were excluded from the dataset, because all plants from population 3040 died and the only surviving plant from the 3072 population was not in a good health to proceed. Leaf length was measured in cm using a ruler from the leaf base to the tip of the longest leaf on each plant and the number of leaves on each plant were counted (Supplementary Tables 4.1 & 4.2).

4.3.4. Statistical Analysis

One-way analysis of variance (ANOVA) was performed in GraphPad Prism 8 to evaluate if differences in the mean germination rates, leaf number and leaf length between parent species and allopolyploids, were statistically significantly ($p < 0.05$). Tukey's test was run following ANOVA to determine if there were any significant comparisons.

4.4. Results

4.4.1. Seed Germination Evaluation

Seeds from certain lineages did not germinate at all and in some cases, either one or two seeds out of three germinated (Supplementary Table 4.3). In total, 96 seeds germinated and were used to evaluate the seed viability (germination percentage) and germination rates. In terms of germination percentages, both short- and long-liguled *T.*

miscellus showed higher percentage of viable seeds compared to diploid parental species (Table 4.2). Among all species, *T. dubius* (41.67%) and short-liguled *T. miscellus* (81.25%) were observed to have lowest and highest germination percentages, respectively (Table 4.2).

Table 4.2: Number of total planted and germinated seeds, as well as germination percentages of examined species.

Species	Total seeds planted	Germinated seeds	Germination percentage (%)
T. dubius	60	25	41.67
T. pratensis	24	15	62.50
T. miscellus (long-liguled)	24	17	70.83
T. miscellus (short-liguled)	48	39	81.25

In the case of germination rates, both short- and long-liguled *T. miscellus* were shown to have slightly faster germination rates compared to diploid parental species (Figure 4.2).

These differences between germination rates of parents and allopolyploid species were statistically insignificant ($p > 0.05$) (Supplementary Table 4.4).

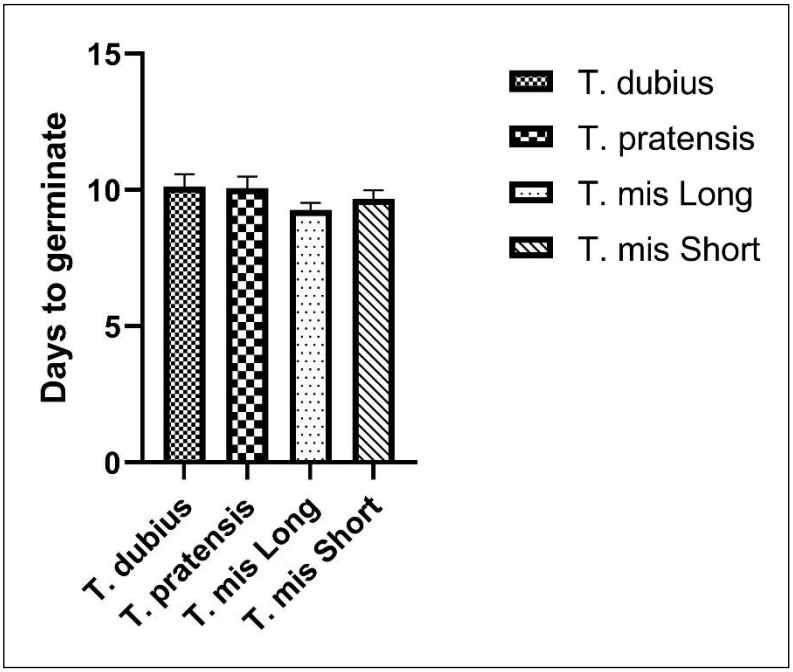


Figure 4.2: Germination rates of *Tragopogon* diploid and allopolyploid species. Each bar represents the mean of all populations from each species.

4.4.2. Morphological Observations

Parent populations were shown to have more leaves as compared to *T. miscellus* allopolyploids (Figure 4.3). However, considerable differences were observed in *T.*

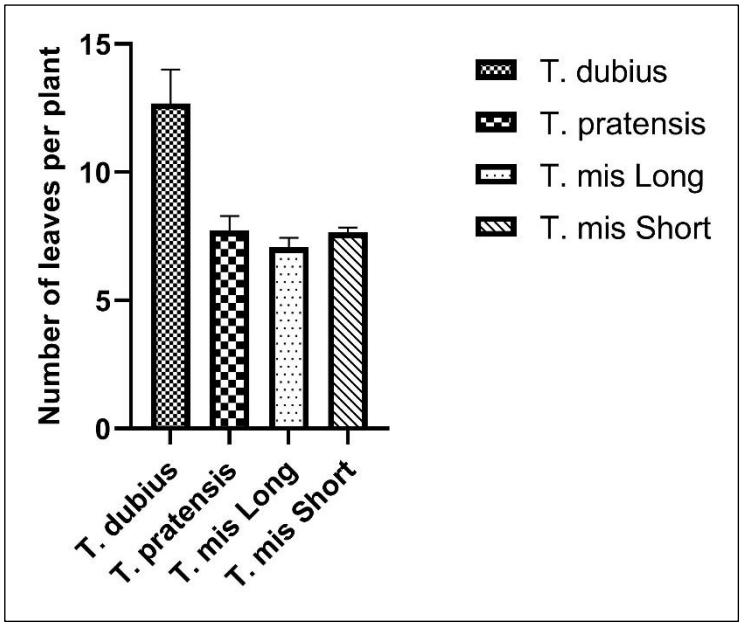


Figure 4.3: Differences between diploid and allotetraploid *Tragopogon* species in terms of number of leaves. Each bar represents mean of all populations from each species.

dubius, which showed significantly ($p < 0.05$) the highest number of leaves compared to both reciprocal allopolyploids (Supplementary Table 4.5). There were insignificant differences observed in leaf number between *T. pratensis* and both short- and long-liguled *T. miscellus* (Supplementary Table 4.5).

Both reciprocal forms of the *T. miscellus* allopolyploid were found to have significantly greater leaf lengths than the diploid parents ($p < 0.05$) (Figure 4.4 & Supplementary Table 4.6). The short-liguled *T. miscellus* showed increased length as compared to long-liguled *T. miscellus* but these differences were insignificant (Supplementary Table 4.6).

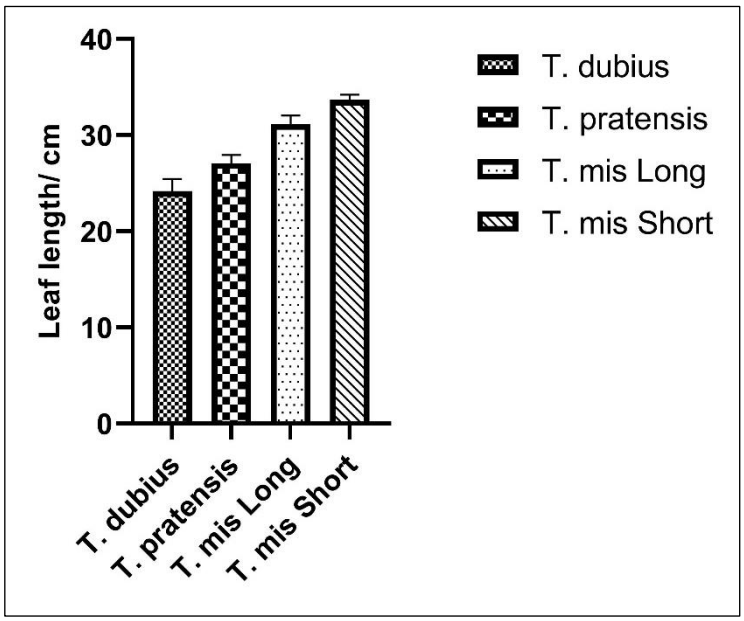


Figure 4.4: Differences between diploid and allotetraploid *Tragopogon* species in terms of leaf length. Each bar represents mean of all populations from each species.

4.5. Discussion

In addition to the genomic, genetic and transcriptomic alterations, polyploidy plays an important role in generating morphological diversity in natural populations of polyploids (Otto and Whitton 2000). Due to the enormous amount of genetic material, the gigas effect including the enlargement of plant cells has been observed in polyploids, which may further lead to larger organs and tissues (Sattler et al. 2016). The gigas effect on phenotypic traits has been associated with the evolutionary advantage of polyploids over their progenitors, facilitating them to successfully establish and survive in challenging environments (Soltis et al. 2014b; Ramsey and Ramsey 2014; Miller et al. 2012). The apparent morphological effects have been noticed in both vegetative and reproductive organs in polyploids (Otto 2007). For instance, polyploids often have larger leaves, flowers and heavier seeds with higher germination rates as well as large stomatal cells compared to diploid parent species (Hoya et al. 2007; Eliášová and Münzbergová

2014; Sun et al. 2015; Van Laere et al. 2011). Thus, understanding the phenotypic consequences of polyploidy is an important aspect in polyploid evolution.

In this study, we have investigated and compared the seed germination rates and percentages of reciprocal forms of *T. miscellus* allopolyploids and their diploid parent species. Our expectation was that *T. miscellus* allopolyploids would have faster germination rates and higher germination percentages than the diploid parent species. The preliminary results confirmed this hypothesis as the examined allopolyploids were generally noticed to have faster germination rates and more viable seeds compared to the diploid parent species. Previous studies have also demonstrated faster germination rates and higher germination percentages of polyploids relative to diploid parent species (Broadhurst et al. 2012; Astuti et al. 2020; Stevens et al. 2020; Selvi and Vivona 2022). The mechanisms behind the faster germination rates of polyploids are not well studied but it is speculated that the extra gene content in polyploid seeds might lead to greater energy reserves and may also require more energy, thus, resulting in rapid mobilisation of seed reserves to aid in growth (Von Well and Fossey 1998). The faster germination rates and higher viability either due to rapid mobilisation of seed reserves, seeds of larger size or other effects might assist polyploids in their successful ecological adaptation (Eliášová and Münzbergová 2014; Stevens et al. 2020).

We also examined and compared the leaf number and length between polyploids and parent species to test if polyploids have more and longer leaves relative to diploids. In terms of leaf number, both diploid parent species, especially *T. dubius* had more leaves compared to the *T. miscellus* allopolyploid forms. On the contrary, a *Cucumis* allopolyploid, *C.* × *hytivus* had significantly more leaves than its diploid parent species (Yu et al. 2021). Similarly, barley tetraploids showed fewer leaves compared to diploids at a given growth time (Chen et al. 2021), which was thought to have occurred due to an

occasionally sustained vegetative growth (Dudits et al. 2016). The trade-off between stomatal cell size and density has been noticed in polyploids, for example, tetraploid barley leaves had a lower density of stomata which were larger in cell size (Chen et al. 2021). It could be possible that a similar trade-off exists between the leaf length and leaf number in *T. miscellus* allopolyploids, but this requires further investigation. In the case of leaf length trait, *T. miscellus* allopolyploids were shown to have longer leaves relative to both parent species. The short-liguled *T. miscellus* even had slightly longer leaves compared to the long-liguled form. However, these observations were statistically insignificant ($p > 0.05$). A past study which conducted a comparison of various morphological traits between the polyploids of *Phlox amabilis* and diploids showed that tetraploids exhibited longer leaves than diploids and hexaploids (Chansler et al. 2016), but hexaploids had thicker leaves among all (Chansler et al. 2016). Likewise, another study on the induced tetraploid *Rhododendron fortunei* showed that the tetraploid plants generally displayed larger and thicker leaves (Lan et al. 2020). However, there were a few plants that showed smaller leaves than diploids, which was attributed to the decrease in the number of cell divisions in the polyploids (Lan et al. 2020). Interestingly, octoploid, *Jatropha curcas* at a higher ploidy level had much smaller leaves than its diploid counterparts but the leaf cells of *J. curcas* were larger (Niu et al. 2016). This suggest that variations in morphological traits at different ploidy levels depend on the plant species being investigated. Interestingly, leaf width of *T. miscellus* allopolyploids is remarkably different than the parent species, especially *T. dubius*. In particular, the short-liguled population 3053 was observed to have the broadest leaves among all. However, this parameter was not measured as a part of this study.

Several potential mechanisms are believed to play a role in the phenotypic changes in polyploids. These mechanisms include genetic and genomic rearrangements,

epigenetic remodelling and transcriptomic alterations (Comai 2005a). For example, homeologous rearrangements might be responsible for influencing changes in flowering time, disease resistance and seed yield in polyploids (Osborn et al. 2003a; Pires et al. 2004; Zhao et al. 2006). For example, the morphological similarity of the allotetraploid *Senecio mohavensis* to its parental species *S. flavus* has been associated with biased loss of homeologs derived from the other parent species, *S. glaucus* (Alexander-Webber et al. 2016). Moreover, the alterations in the expression of light-harvesting complex genes of photosystem II were also observed between the allotetraploids, *Dactylorhiza traunsteineri* and *D. majalis* which correspond to the morphological and physiological changes in these allopolyploids, such as chlorophyll content and photosynthetic rates (Wolfe et al. 2021). In our expression work given in chapter 3, there were a few differentially expressed organelle-targeted nuclear genes in leaf tissues between each parent species and *T. miscellus* allopolyploids as well as between reciprocal forms. In future, these genes could be targeted with respect to the leaf morphological and physiological parameters to investigate their effect on the phenotypic variations in *T. miscellus* allopolyploids. The correlation between the genetic/transcriptomic changes and phenotypic responses in polyploids could be a potential way to investigate the integrated role of genes and phenotypic traits. Such genes could later be exploited for crop improvement and developing plant species with desirable traits.

Chapter 5

Conclusions and Future Directions

This study was conducted to investigate the evolution of cytonuclear interactions in reciprocally formed *Tragopogon miscellus* allopolyploids. In this regard, I have primarily examined the transcriptomic consequences of cytonuclear genes implicated in chimeric enzyme complexes, and dual-targeted genes in the short- and long-liguled *T. miscellus*. Gene expression analyses revealed that the hybridization and whole genome duplication events in *T. miscellus* allopolyploids led to differential regulation of expression at the homeolog level instead of the whole gene expression level (sum of both homeologs), as both short- and long-liguled *T. miscellus* allopolyploids showed total gene expression levels similar to their parental species. The expression regulation at the homeolog level was evidenced by novel expression bias patterns that were not exhibited by corresponding loci in both parent species, and by retaining the vertically inherited expression differences of their parents. Such differential regulation of homeologs in *T. miscellus* allopolyploids might suggest the occurrence of either epigenetic or post-transcriptional/translational regulation. This may involve differences in cis-/trans-regulatory elements, TE density, alternative splicing factors/marks and translational machinery between homeologs, hence resulting in differential expression regulation. Therefore, future studies investigating the role of these factors or mechanisms at epigenetic and post-transcriptional/translational levels in *T. miscellus* allopolyploids could explain how organelle-targeted duplicated nuclear genes are altering their expression patterns. Proteomic data for *T. miscellus* allopolyploids and parents were collected to investigate if cytonuclear genes were employing adjustments at protein levels. Unfortunately, because of severe delays due to COVID the results of this analysis are not included in the thesis. These data will be included in another future publication.

Interestingly, the transcriptomic analysis also revealed the unbalanced homeolog expression bias toward *T. pratensis* in both reciprocal forms of *T. miscellus* allopolyploids. This might indicate that *T. pratensis* may have advantageous features including lower TE density and small RNA targeting sites etc., leading to the expression preference of *T. pratensis* derived homeologs in both reciprocal forms. For example, preliminary genome sequencing data (unpublished) of both parent species, *T. dubius* and *T. pratensis*, suggest that the *T. dubius* genome may have more transposons compared to *T. pratensis*. These transposons may be influencing the expression of nearby genes through silencing via epigenetic mechanisms such as DNA methylation, histone modifications, etc., thereby downregulating the expression of *T. dubius* homeologs. However, further validation with respect to differentially expressed cytonuclear genes is needed to examine if such differences in transposon density exist between parents and are subsequently inherited in polyploids, resulting in the differential regulation of homeologs.

Organelle gene expression was not up-regulated in *T. miscellus* allopolyploids, contrary to expectations that organelle genes will coordinate their expression with the duplicated nuclear genes through up-regulated expression. Because total gene expression of duplicated nuclear genes was similar to the parental expression levels in both reciprocal forms, there may not be the need to up-regulate organelle expression. As *T. miscellus* allopolyploids are young (90-100 years), they may need additional evolutionary time to allow them to establish specific patterns or incorporate changes at total gene expression levels. Hence, given time, they may develop particular patterns, which may eventually demand expression changes in organellar genes. Thus, tracking expression changes in later generations of *T. miscellus* allopolyploids would help us to confirm this case.

In terms of morphological responses in *T. miscellus* allopolyploids, we have noticed faster germination rates and longer leaves in allopolyploids compared to their

diploid parent species. Relating these morphological traits to the physiological processes including photosynthesis, chlorophyll content, seed structure and seedling growth, etc., and to the interactions among cytonuclear genes would be helpful in understanding the ecological success and evolution of allopolyploids.

References

Abdel-Ghany SE, Hamilton M, Jacobi JL, Ngam P, Devitt N, Schilkey F, Ben-Hur A, Reddy AS (2016) A survey of the *Sorghum* transcriptome using single-molecule long reads. Nature communications 7:11706.

Adams KL (2007) Evolution of duplicate gene expression in polyploid and hybrid plants. Journal of heredity 98 (2):136-141.

Adams KL, Cronn R, Percifield R, Wendel JF (2003a) Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proceedings of the national academy of sciences 100 (8):4649-4654.

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 100 (8):4649-4654.

Adams KL, Daley DO, Qiu Y-L, Whelan J, Palmer JD (2000) Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. Nature 408 (6810):354-357.

Adams KL, Palmer JD (2003) Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Molecular phylogenetics and evolution 29 (3):380-395.

Adams KL, Percifield R, Wendel JF (2004) Organ-specific silencing of duplicated genes in a newly synthesized cotton allotetraploid. Genetics 168 (4):2217-2226.

Adams KL, Qiu Y-L, Stoutemyer M, Palmer JD (2002) Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. Proceedings of the national academy of sciences 99 (15):9905-9912.

Adams KL, Rosenblueth M, Qiu Y-L, Palmer JD (2001) Multiple losses and transfers to the nucleus of two mitochondrial succinate dehydrogenase genes during angiosperm evolution. Genetics 158 (3):1289-1300.

Adams KL, Wendel JF (2005) Polyploidy and genome evolution in plants. Current opinion in plant biology 8 (2):135-141.

Ahloowalia B, Garber E (1961) The genus *Collinsia*. XIII. Cytogenetic studies of interspecific hybrids involving species with pediceled flowers. Botanical gazette 122 (3):219-228.

Ainouche M, Chelaifa H, Ferreira J, Bellot S, Aïnouche A, Salmon A (2012) Erratum from—polyploid evolution in *Spartina*: dealing with highly redundant hybrid genomes. In: Polyploidy and genome evolution. Springer, pp 225-243.

Ainouche ML, Baumel A, Salmon A, Yannic G (2004) Hybridization, polyploidy and speciation in *Spartina* (Poaceae). New Phytologist 161 (1):165-172.

Albertin W, Alix K, Balliau T, Brabant P, Davanture M, Malosse C, Valot B, Thiellement H (2007) Differential regulation of gene products in newly synthesized *Brassica napus* allotetraploids is not related to protein function nor subcellular localization. BMC genomics 8 (1):56.

Albertin W, Balliau T, Brabant P, Chevre A-M, Eber F, Malosse C, Thiellement H (2006) Numerous and rapid nonstochastic modifications of gene products in newly synthesized *Brassica napus* allotetraploids. Genetics 173 (2):1101-1113.

Alexander-Webber D, Abbott RJ, Chapman MA (2016) Morphological convergence between an allopolyploid and one of its parental species correlates with biased gene expression and DNA Loss. Journal of heredity 107 (5):445-454.

Amiri S, Kazemitabaar S, Ranjbar G, Azadbakht M (2010) The effect of trifluralin and colchicine treatments on morphological characteristics of jimsonweed (*Datura Stramonium L.*). Trakia journal of sciences 8 (4):47-61.

Andersson MX, Goksör M, Sandelius AS (2007) Optical manipulation reveals strong attracting forces at membrane contact sites between endoplasmic reticulum and chloroplasts. Journal of Biological Chemistry 282 (2):1170-1174.

Anssour S, Krügel T, Sharbel T, Saluz H, Bonaventure G, Baldwin I (2009) Phenotypic, genetic and genomic consequences of natural and synthetic polyploidization of *Nicotiana attenuata* and *Nicotiana obtusifolia*. Annals of Botany 103 (8):1207-1217.

Arias S, Bhatia SK (2015) Barley. In: Medical applications for biomaterials in Bolivia. Springer, pp 15-22.

Astuti G, Pratesi S, Peruzzi L, Carta A (2020) Two closely related *Tulipa* species with different ploidy levels show distinct germination rates. Seed Science Research 30 (1):45-48.

Auger B, Baron C, Lucas M-O, Vautrin S, Bergès H, Chalhoub B, Fautrel A, Renard M, Nesi N (2009) *Brassica* orthologs from BANYULS belong to a small multigene family, which is involved in procyanidin accumulation in the seed. Planta 230 (6):1167.

Austin R, Morgan C, Ford M, Bhagwat S (1982) Flag leaf photosynthesis of *Triticum aestivum* and related diploid and tetraploid species. Annals of Botany 49 (2):177-189.

Axtell MJ (2013) Classification and comparison of small RNAs from plants. Annual review of plant biology 64:137-159.

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 (1):256-265.

Baniaga AE, Marx HE, Arrigo N, Barker MS (2020) Polyploid plants have faster rates of multivariate niche differentiation than their diploid relatives. Ecology letters 23 (1):68-78.

Bao Y, Hu GJ, Grover CE, Conover J, Yuan DJ, Wendel JF (2019) Unraveling cis and trans regulatory evolution during cotton domestication. Nature communications 10.

Bardil A, de Almeida JD, Combes MC, Lashermes P, Bertrand B (2011) Genomic expression dominance in the natural allopolyploid Coffea arabica is massively affected by growth temperature. New Phytologist 192 (3):760-774.

Barker MS, Arrigo N, Baniaga AE, Li Z, Levin DA (2016) On the relative abundance of autopolyploids and allopolyploids. New Phytologist 210 (2):391-398.

Barker MS, Kane NC, Matvienko M, Kozik A, Michelmore RW, Knapp SJ, Rieseberg LH (2008) Multiple paleopolyploidizations during the evolution of the Compositae reveal parallel patterns of duplicate gene retention after millions of years. Molecular Biology and Evolution 25 (11):2445-2455.

Beaulieu JM, Leitch IJ, Patel S, Pendharkar A, Knight CA (2008) Genome size is a strong predictor of cell size and stomatal density in angiosperms. New Phytologist 179 (4):975-986.

Bell GD, Kane NC, Rieseberg LH, Adams KL (2013) RNA-seq analysis of allele-specific expression, hybrid effects, and regulatory divergence in hybrids compared with their parents from natural populations. Genome biology and evolution 5 (7):1309-1323.

Bilgrami S, Houshmand S, Khodambashi M, Zandi P, Siavoshi M, Khademi S, Navabpour S, Dehbaneh MN, Amouei H, Tadayyon M (2015) Photosynthetic performance in ploidy levels and amphyploids of wheat during developmental stages. Journal of animal & plant sciences 25 (6).

Birchler JA, Veitia RA (2007) The gene balance hypothesis: from classical genetics to modern genomics. The Plant Cell 19 (2):395-402.

Birchler JA, Veitia RA (2010) The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. New Phytologist 186 (1):54-62.

Birchler JA, Veitia RA (2012) Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. Proceedings of the national Academy of sciences 109 (37):14746-14753.

Blanc G, Wolfe KH (2004) Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. The Plant Cell 16 (7):1679-1691.

Blencowe BJ (2006) Alternative splicing: new insights from global analyses. Cell 126 (1):37-47.

Boatwright JL, McIntyre LM, Morse AM, Chen S, Yoo M-J, Koh J, Soltis PS, Soltis DE, Barbazuk WB (2018a) A robust methodology for assessing differential homeolog contributions to the transcriptomes of allopolyploids. Genetics 210 (3):883-894.

Boatwright JL, McIntyre LM, Morse AM, Chen SX, Yoo MJ, Koh J, Soltis PS, Soltis DE, Barbazuk WB (2018b) A Robust Methodology for Assessing Differential Homeolog Contributions to the Transcriptomes of Allopolyploids. Genetics 210 (3):883-894.
Bock DG, Andrew RL, Rieseberg LH (2014) On the adaptive value of cytoplasmic genomes in plants. Molecular ecology 23 (20):4899-4911.
Bomblies K (2020) When everything changes at once: finding a new normal after genome duplication. Proceedings of the Royal Society B 287 (1939):20202154.
Bottley A, Xia G, Koebner R (2006) Homoeologous gene silencing in hexaploid wheat. The Plant Journal 47 (6):897-906.
Bourc'his D, Voinnet O (2010) A small-RNA perspective on gametogenesis, fertilization, and early zygotic development. Science 330 (6004):617-622.
Brehm BG, Ownbey M (1965) Variation in chromatographic patterns in the *Tragopogon dubius-pratensis-porrifolius* complex (Compositae). American journal of botany 52 (8):811-818.
Bretagnolle Fa, Thompson JD (1995) Gametes with the somatic chromosome number: mechanisms of their formation and role in the evolution of autopolyploid plants. New Phytologist 129 (1):1-22.
Broadhurst LM, Murray BG, Forrester R, Young AG (2012) Cryptic genetic variability in *Swainsona sericea* (A. Lee) *H. Eichler* (Fabaceae): lessons for restoration. Australian Journal of Botany 60 (5):429-438.
Brownfield L, Köhler C (2010) Unreduced gamete formation in plants: mechanisms and prospects. Journal of experimental botany 62 (5):1659-1668.
Bruce BD (2000) Chloroplast transit peptides: structure, function and evolution. Trends in cell biology 10 (10):440-447.
Buggs R, Doust A, Tate J, Koh J, Soltis K, Feltus F, Paterson A, Soltis P, Soltis D (2009a) Gene loss and silencing in Tragopogon miscellus (Asteraceae): comparison of natural and synthetic allotetraploids. Heredity 103 (1):73-81.
Buggs R, Doust A, Tate J, Koh J, Soltis K, Feltus F, Paterson A, Soltis P, Soltis D (2009b) Gene loss and silencing in *Tragopogon miscellus* (Asteraceae): comparison of natural and synthetic allotetraploids. Heredity 103 (1):73.
Buggs RJ, Chamala S, Wu W, Gao L, May GD, Schnable PS, Soltis DE, Soltis PS, Barbazuk WB (2010a) 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 RJ, Chamala S, Wu W, Tate JA, Schnable PS, Soltis DE, Soltis PS, Barbazuk WB (2012a) Rapid, repeated, and clustered loss of duplicate genes in allopolyploid plant populations of independent origin. Current biology 22 (3):248-252.
Buggs RJ, Elliott NM, Zhang L, Koh J, Viccini LF, Soltis DE, Soltis PS (2010b) Tissue-specific silencing of homoeologs in natural populations of the recent allopolyploid *Tragopogon mirus*. New Phytologist 186 (1):175-183.
Buggs RJ, Renny-Byfield S, Chester M, Jordon-Thaden IE, Viccini LF, Chamala S, Leitch AR, Schnable PS, Barbazuk WB, Soltis PS (2012b) Next-generation sequencing and genome evolution in allopolyploids. American journal of botany 99 (2):372-382.
Buggs RJ, Wendel JF, Doyle JJ, Soltis DE, Soltis PS, Coate JE (2014) The legacy of diploid progenitors in allopolyploid gene expression patterns. Philosophical Transactions of the Royal Society B: Biological Sciences 369 (1648):20130354.
Buggs RJ, Zhang L, Miles N, Tate JA, Gao L, Wei W, Schnable PS, Barbazuk WB, Soltis PS, Soltis DE (2011a) Transcriptomic shock generates evolutionary novelty in a newly formed, natural allopolyploid plant. Current biology 21 (7):551-556.
Buggs RJA, Chamala S, Wu W, Gao L, May GD, Schnable PS, Soltis DE, Soltis PS, Barbazuk WB (2010c) 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, Zhang LJ, Miles N, Tate JA, Gao L, Wei W, Schnable PS, Barbazuk WB, Soltis PS, Soltis DE (2011b) Transcriptomic Shock Generates Evolutionary Novelty in a Newly Formed, Natural Allopolyploid Plant. Current biology 21 (7):551-556.

Burton RS, Pereira RJ, Barreto FS (2013) Cytonuclear genomic interactions and hybrid breakdown. Annual review of ecology, evolution, and systematics 44 (1):281-302.
Busch A, Hertel KJ (2012) Evolution of SR protein and hnRNP splicing regulatory factors. Wiley Interdisciplinary Reviews: RNA 3 (1):1-12.
Bush SJ, Chen L, Tovar-Corona JM, Urrutia AO (2017) Alternative splicing and the evolution of phenotypic novelty. Philosophical Transactions of the Royal Society B: Biological Sciences 372 (1713):20150474.
Buttgereit F, Brand MD (1995) A hierarchy of ATP-consuming processes in mammalian cells. Biochemical Journal 312 (1):163-167.
Campbell MA, Haas BJ, Hamilton JP, Mount SM, Buell CR (2006) Comprehensive analysis of alternative splicing in rice and comparative analyses with *Arabidopsis*. BMC genomics 7 (1):327.
Caperta A, Delgado M, Ressurreição F, Meister A, Jones RN, Viegas W, Houben A (2006) Colchicine-induced polyploidization depends on tubulin polymerization in c-metaphase cells. Protoplasma 227 (2-4):147-153.
Cardol P, Vanrobaeys F, Devreese B, Van Beeumen J, Matagne RF, Remacle C (2004) Higher plant-like subunit composition of mitochondrial complex I from *Chlamydomonas reinhardtii*: 31 conserved components among eukaryotes. Biochimica et Biophysica Acta (BBA)-Bioenergetics 1658 (3):212-224.
Carputo D, Basile B, Cardi T, Frusciante L (2000) Erwinia resistance in backcross progenies of *Solanum tuberosum x S. tarijense* and *S. tuberosum* (+) *S. commersonii* hybrids. Potato research 43 (2):135-142.
Carrie C, Small I (2013) A reevaluation of dual-targeting of proteins to mitochondria and chloroplasts. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research 1833 (2):253-259.
Carrie C, Whelan J (2013) Widespread dual targeting of proteins in land plants: when, where, how and why. Plant signaling & behavior 8 (8):644-662.
Carroll J, Fearnley IM, Skehel JM, Shannon RJ, Hirst J, Walker JE (2006) Bovine complex I is a complex of 45 different subunits. Journal of Biological Chemistry 281 (43):32724-32727.
Cenci A, Combes M-C, Lashermes P (2012) Genome evolution in diploid and tetraploid *Coffea* species as revealed by comparative analysis of orthologous genome segments. Plant molecular biology 78 (1-2):135-145.
Chagué V, Just J, Mestiri I, Balzergue S, Tanguy AM, Huneau C, Huteau V, Belcram H, Coriton O, Jahier J (2010) Genome-wide gene expression changes in genetically stable synthetic and natural wheat allohexaploids. New Phytologist 187 (4):1181-1194.
Chalhoub B, Denoeud F, Liu S, Parkin IA, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B (2014) Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. Science 345 (6199):950-953.
Chang PL, Dilkes BP, McMahon M, Comai L, Nuzhdin SV (2010) Homoeolog-specific retention and use in allotetraploid *Arabidopsis suecica* depends on parent of origin and network partners. Genome biology 11 (12):R125.
Chansler MT, Ferguson CJ, Fehlberg SD, Prather LA (2016) The role of polyploidy in shaping morphological diversity in natural populations of *Phlox amabilis*. American journal of botany 103 (9):1546-1558.
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 (2):503-517.
Chelaifa H, Monnier A, Ainouche M (2010) Transcriptomic changes following recent natural hybridization and allopolyploidy in the salt marsh species *Spartina× townsendii* and *Spartina anglica* (Poaceae). New Phytologist 186 (1):161-174.
Chen G, SUN WB, Sun H (2009) Morphological characteristics of leaf epidermis and size variation of leaf, flower and fruit in different ploidy levels in *Buddleja macrostachya* (Buddlejaceae). Journal of Systematics and Evolution 47 (3):231-236.
Chen S, Zhou Y, Chen Y, Gu J (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34 (17):i884-i890.

Chen Y, Xu H, He T, Gao R, Guo G, Lu R, Chen Z, Liu C (2021) Comparative Analysis of Morphology, Photosynthetic Physiology, and Transcriptome Between Diploid and Tetraploid Barley Derived From Microspore Culture. Frontiers in plant science 12:626916.

Chen ZJ (2007a) Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. Annual review of plant biology 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.

Chen ZJ (2010) Molecular mechanisms of polyploidy and hybrid vigor. Trends in plant science 15 (2):57-71.

Chen ZJ, Ni Z (2006a) Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. Bioessays 28 (3):240-252.

Chen ZJ, Ni ZJB (2006b) Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. 28 (3):240-252.

Chen ZJ, Tian L (2007) Roles of dynamic and reversible histone acetylation in plant development and polyploidy. Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression 1769 (5-6):295-307.

Cheng F, Sun C, Wu J, Schnable J, Woodhouse MR, Liang J, Cai C, Freeling M, Wang X (2016) Epigenetic regulation of subgenome dominance following whole genome triplication in *Brassica rapa*. New Phytologist 211 (1):288-299.

Cheng F, Wu J, Fang L, Sun S, Liu B, Lin K, Bonnema G, Wang X (2012) Biased gene fractionation and dominant gene expression among the subgenomes of *Brassica rapa*. PloS one 7 (5):e36442.

Chester M, Gallagher JP, Symonds VV, Cruz da Silva AV, Mavrodiev EV, Leitch AR, Soltis PS, Soltis DE (2012a) Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon miscellus* (Asteraceae). Proceedings of the national Academy of sciences 109 (4):1176-1181.

Chester M, Gallagher JP, Symonds VV, da Silva AVC, Mavrodiev EV, Leitch AR, Soltis PS, Soltis DE (2012b) Extensive chromosomal variation in a recently formed natural allopolyploid species, Tragopogon miscellus (Asteraceae). Proceedings of the national Academy of sciences 109 (4):1176-1181.

Chester M, Riley R, Soltis P, Soltis D (2015) Patterns of chromosomal variation in natural populations of the neoallotetraploid *Tragopogon mirus* (Asteraceae). Heredity 114 (3):309.

Chew O, Whelan J, Millar AH (2003) Molecular definition of the ascorbate-glutathione cycle in *Arabidopsis* mitochondria reveals dual targeting of antioxidant defenses in plants. Journal of Biological Chemistry 278 (47):46869-46877.

Chuong EB, Elde NC, Feschotte C (2017) Regulatory activities of transposable elements: from conflicts to benefits. Nature Reviews Genetics 18 (2):71.

Coate JE, Bar H, Doyle JJ (2014) Extensive translational regulation of gene expression in an allopolyploid (*Glycine dolichocarpa*). The Plant Cell 26 (1):136-150.

Coate JE, Schreyer WM, Kum D, Doyle JJ (2020) Robust cytonuclear coordination of transcription in nascent *Arabidopsis thaliana* autopolyploids. Genes 11 (2).

Coleman L (1950) Nuclear conditions in normal stem tissue of *Vicia faba*. Canadian Journal of Research 28 (3):382-391.

Comai L (2005a) The advantages and disadvantages of being polyploid. Nature reviews genetics 6 (11):836-846.

Comai L (2005b) The advantages and disadvantages of being polyploid. Nature Reviews Genetics 6 (11):836.

Combes M-C, Cenci A, Baraille H, Bertrand B, Lashermes P (2011) Homeologous gene expression in response to growing temperature in a recent allopolyploid (*Coffea arabica L.*). Journal of heredity 103 (1):36-46.

Combes MC, Dereeper A, Severac D, Bertrand B, Lashermes P (2013) Contribution of subgenomes to the transcriptome and their intertwined regulation in the allopolyploid Coffea arabica grown at contrasted temperatures. New Phytologist 200 (1):251-260.

Conant GC, Birchler JA, Pires JC (2014) Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time. Current opinion in plant biology 19:91-98.
Conant GC, Wolfe KH (2008) Turning a hobby into a job: how duplicated genes find new functions. Nature Reviews Genetics 9 (12):938.
Coyne J, Orr H (2004) Speciation. 2004, Sunderland, MA. Sinauer Associates, Inc.
Creissen G, Reynolds H, Xue Y, Mullineaux P (1995) Simultaneous targeting of pea glutathione reductase and of a bacterial fusion protein to chloroplasts and mitochondria in transgenic tobacco. The Plant Journal 8 (2):167-175.
Cui L, Wall PK, Leebens-Mack JH, Lindsay BG, Soltis DE, Doyle JJ, Soltis PS, Carlson JE, Arumuganathan K, Barakat A (2006) Widespread genome duplications throughout the history of flowering plants. Genome research 16 (6):738-749.
D'Amato F (1952) Polyploidy in the differentiation and function of tissues and cells in plants: a critical examination of the literature. Caryologia 4 (3):311-358.
D'Amato F (1964) Endopolyploidy as a factor in plant tissue development. Caryologia 17 (1):41-52.
d'Erfurth I, Jolivet S, Froger N, Catrice O, Novatchkova M, Simon M, Jenczewski E, Mercier R (2008) Mutations in AtPS1 (*Arabidopsis thaliana* parallel spindle 1) lead to the production of diploid pollen grains. PLoS genetics 4 (11):e1000274.
De Bodt S, Maere S, Van de Peer Y (2005) Genome duplication and the origin of angiosperms. Trends in ecology & evolution 20 (11):591-597.
De Smet R, Adams KL, Vandepoele K, Van Montagu MC, Maere S, Van de Peer Y (2013a) Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. Proceedings of the national Academy of sciences 110 (8):2898-2903.
De Smet R, Adams KL, Vandepoele K, van Montagu MCE, Maere S, Van de Peer Y (2013b) Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. Proceedings of the National Academy of Sciences of the United States of America 110 (8):2898-2903.
De Storme N, Geelen D (2013) Sexual polyploidization in plants–cytological mechanisms and molecular regulation. New Phytologist 198 (3):670-684.
Diaz De La Guardia C, Blanca G (2004) A new Spanish species of *Tragopogon* (Asteraceae: Lactuceae). Botanical journal of the Linnean Society 146 (4):505-511.
Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2012) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29 (1):15-21.
Dong S, Adams KL (2011) Differential contributions to the transcriptome of duplicated genes in response to abiotic stresses in natural and synthetic polyploids. New Phytologist 190 (4):1045-1057.
Doolittle WF, Sapienza C (1980) Selfish genes, the phenotype paradigm and genome evolution. Nature 284 (5757):601.
Doyle JJ, Coate JE (2019) Polyploidy, the nucleotype, and novelty: the impact of genome doubling on the biology of the cell. International Journal of Plant Sciences 180 (1):1-52.
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.
Doyle JJ, Sherman-Broyles S (2017) Double trouble: taxonomy and definitions of polyploidy. New Phytologist 213 (2):487-493.
Drechsel G, Kahles A, Kesarwani AK, Stauffer E, Behr J, Drewe P, Rätsch G, Wachter A (2013) Nonsense-mediated decay of alternative precursor mRNA splicing variants is a major determinant of the *Arabidopsis* steady state transcriptome. The Plant Cell 25 (10):3726-3742.
Duchemin W, Dupont P-Y, Campbell MA, Ganley ARD, Cox MP (2015) HyLiTE: accurate and flexible analysis of gene expression in hybrid and allopolyploid species. BMC bioinformatics 16 (1):8.

Dudits D, Török K, Cseri A, Paul K, Nagy AV, Nagy B, Sass L, Ferenc G, Vankova R, Dobrev P (2016) Response of organ structure and physiology to autotetraploidization in early development of energy willow *Salix viminalis*. Plant physiology 170 (3):1504-1523.

Edger PP, Pires JC (2009) Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. Chromosome Research 17 (5):699.

Edger PP, Smith R, McKain MR, Cooley AM, Vallejo-Marin M, Yuan Y, Bewick AJ, Ji L, Platts AE, Bowman MJ (2017) Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. The Plant Cell 29 (9):2150-2167.

Eichten SR, Ellis NA, Makarevitch I, Yeh C-T, Gent JI, Guo L, McGinnis KM, Zhang X, Schnable PS, Vaughn MW (2012) Spreading of heterochromatin is limited to specific families of maize retrotransposons. PLoS genetics 8 (12):e1003127.

Elias E, Steiger D, Cantrell R (1996) Evaluation of lines derived from wild emmer chromosome substitutions: II. Agronomic traits. Crop Science 36 (2):228-233.

Eliášová A, Münzbergová Z (2014) Higher seed size and germination rate may favour autotetraploids of *Vicia cracca L.*(Fabaceae). Biological Journal of the Linnean Society 113 (1):57-73.

Expósito-Rodríguez M, Borges AA, Borges-Pérez A, Pérez JA (2008) Selection of internal control genes for quantitative real-time RT-PCR studies during tomato development process. BMC Plant Biology 8 (1):1-12.

Fahima T, Cheng J, Peng J, Nevo E, Korol A Asymmetry distribution of disease resistance genes and domestication synrome QTLs in tetraploid wheat genome. In: 8th International Congress of Plant Molecular Biology, Adelaide, Australia, 2006.

Fawcett JA, Maere S, Van De Peer Y (2009) Plants with double genomes might have had a better chance to survive the Cretaceous–Tertiary extinction event. Proceedings of the national Academy of sciences 106 (14):5737-5742.

Fei Q, Xia R, Meyers BC (2013) Phased, secondary, small interfering RNAs in posttranscriptional regulatory networks. The Plant Cell 25 (7):2400-2415.

Feldman M, Levy AA, Fahima T, Korol A (2012) Genomic asymmetry in allopolyploid plants: wheat as a model. Journal of experimental botany 63 (14):5045-5059.

Feng J, Li J, Gao Z, Lu Y, Yu J, Zheng Q, Yan S, Zhang W, He H, Ma L (2015) SKIP confers osmotic tolerance during salt stress by controlling alternative gene splicing in *Arabidopsis*. Molecular plant 8 (7):1038-1052.

Fernandes Gyorfy M, Miller ER, Conover JL, Grover CE, Wendel JF, Sloan DB, Sharbrough J (2021) Nuclear-cytoplasmic balance: whole genome duplications induce elevated organellar genome copy number. The Plant Journal.

Ferrando A, Castellano M, Lisón P, Leister D, Stepanova AN, Hanson J (2017) Relevance of translational regulation on plant growth and environmental responses. Frontiers in plant science 8:2170.

Ferreira de Carvalho J, Lucas J, Deniot G, Falentin C, Filangi O, Gilet M, Legeai F, Lode M, Morice J, Trotoux G (2019) Cytonuclear interactions remain stable during allopolyploid evolution despite repeated whole-genome duplications in *Brassica*. The Plant Journal 98 (3):434-447.

Filichkin S, Priest HD, Megraw M, Mockler TC (2015) Alternative splicing in plants: directing traffic at the crossroads of adaptation and environmental stress. Current opinion in plant biology 24:125-135.

Filichkin SA, Priest HD, Givan SA, Shen R, Bryant DW, Fox SE, Wong W-K, Mockler TC (2010) Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. Genome research 20 (1):45-58.

Fishman L, Sweigart AL (2018) When two rights make a wrong: the evolutionary genetics of plant hybrid incompatibilities. Annual review of plant biology 69:707-731.

Fishman L, Willis JH (2006) A cytonuclear incompatibility causes anther sterility in *Mimulus* hybrids. Evolution 60 (7):1372-1381.

Flagel L, Udall J, Nettleton D, Wendel J (2008) Duplicate gene expression in allopolyploid *Gossypium* reveals two temporally distinct phases of expression evolution. BMC biology 6 (1):16.

Flagel LE, Wendel JF (2010) Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation. New Phytologist 186 (1):184-193.

Forsythe ES, Sharbrough J, Havird JC, Warren JM, Sloan DB (2019) CyMIRA: The cytonuclear molecular interactions reference for *Arabidopsis*. Genome biology and evolution 11 (8):2194-2202.

Freeling M (2009) Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. Annual review of plant biology 60:433-453.

Freeling M, Scanlon MJ, Fowler JE (2015) Fractionation and subfunctionalization following genome duplications: mechanisms that drive gene content and their consequences. Current opinion in genetics & development 35:110-118.

Freeling M, Woodhouse MR, Subramaniam S, Turco G, Lisch D, Schnable JC (2012) Fractionation mutagenesis and similar consequences of mechanisms removing dispensable or less-expressed DNA in plants. Current opinion in plant biology 15 (2):131-139.

Friedli M, Trono D (2015) The developmental control of transposable elements and the evolution of higher species. Annual review of cell and developmental biology 31:429-451.

Fu X-D, Ares Jr M (2014) Context-dependent control of alternative splicing by RNA-binding proteins. Nature Reviews Genetics 15 (10):689.

Gaeta RT, Pires JC (2010) Homoeologous recombination in allopolyploids: the polyploid ratchet. New Phytologist 186 (1):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. The Plant Cell 19 (11):3403-3417.

Galichet A, Hoyerová K, Kamínek M, Gruissem W (2008) Farnesylation directs AtIPT3 subcellular localization and modulates cytokinin biosynthesis in *Arabidopsis*. Plant physiology 146 (3):1155-1164.

Galindo-González L, Sarmiento F, Quimbaya M (2018) Shaping plant adaptability, genome structure and gene expression through transposable element epigenetic control: Focus on methylation. Agronomy 8 (9):180.

Gao S, Yan Q, Chen L, Song Y, Li J, Fu C, Dong M (2017) Effects of ploidy level and haplotype on variation of photosynthetic traits: Novel evidence from two *Fragaria* species. PloS one 12 (6):e0179899.

Germain A, Hotto AM, Barkan A, Stern DB (2013) RNA processing and decay in plastids. Wiley Interdisciplinary Reviews: RNA 4 (3):295-316.

Giegé P, Duchêne A-M (2012) Dual localized mitochondrial and nuclear proteins as gene expression regulators in plants? Frontiers in plant science 3:221.

Giegé P, Sweetlove LJ, Cognat V, Leaver CJ (2005) Coordination of nuclear and mitochondrial genome expression during mitochondrial biogenesis in *Arabidopsis*. The Plant Cell 17 (5):1497-1512.

Gifford WD, Pfaff SL, Macfarlan TS (2013) Transposable elements as genetic regulatory substrates in early development. Trends in cell biology 23 (5):218-226.

Giraud E, Ng S, Carrie C, Duncan O, Low J, Lee CP, Van Aken O, Millar AH, Murcha M, Whelan J (2010) TCP transcription factors link the regulation of genes encoding mitochondrial proteins with the circadian clock in *Arabidopsis thaliana*. The Plant Cell 22 (12):3921-3934.

Gong L, Olson M, Wendel JF (2014) Cytonuclear evolution of rubisco in four allopolyploid lineages. Molecular Biology and Evolution 31 (10):2624-2636.

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 (10):3023-3036.

Gottlieb L (2003) Plant polyploidy: gene expression and genetic redundancy. Nature Publishing Group.
Gout J-F, Johri P, Arnaiz O, Doak TG, Bhullar S, Couloux A, Guerin F, Malinsky S, Sperling L, Labadie K (2019) Universal trends of post-duplication evolution revealed by the genomes of 13 Paramecium species sharing an ancestral whole-genome duplication. bioRxiv:573576.
Grant V (1978) Genetics of flowering plants. Columbia University Press.
Grant V (1981a) Plant speciation. In: Plant Speciation. Columbia university press.
Grant V (1981b) Plant speciation. Columbia University Press.
Graveley BR (2001) Alternative splicing: increasing diversity in the proteomic world. Trends in Genetics 17 (2):100-107.
Gray M, Archibald J (2012a) Genomics of chloroplasts and mitochondria. Advances in photosynthesis and respiration.
Gray MW, Archibald JM (2012b) Origins of mitochondria and plastids. In: Genomics of chloroplasts and mitochondria. Springer, pp 1-30.
Greiner S, Bock R (2013) Tuning a ménage à trois: co-evolution and co-adaptation of nuclear and organellar genomes in plants. Bioessays 35 (4):354-365.
Greiner S, Rauwolf U, Meurer J, Herrmann RG (2011) The role of plastids in plant speciation. Molecular ecology 20 (4):671-691.
Greiner S, Wang X, Herrmann RG, Rauwolf U, Mayer K, Haberer G, Meurer J (2008) The complete nucleotide sequences of the 5 genetically distinct plastid genomes of Oenothera, subsection Oenothera: II. A microevolutionary view using bioinformatics and formal genetic data. Molecular Biology and Evolution 25 (9):2019-2030.
Griffiths S, Sharp R, Foote TN, Bertin I, Wanous M, Reader S, Colas I, Moore G (2006) Molecular characterization of Ph1 as a major chromosome pairing locus in polyploid wheat. Nature 439 (7077):749.
Grover C, Gallagher J, Szadkowski E, Yoo M, Flagel L, Wendel J (2012a) Homoeolog expression bias and expression level dominance in allopolyploids. New Phytologist 196 (4):966-971.
Grover CE, Forsythe ES, Sharbrough J, Miller ER, Conover JL, DeTar RA, Chavarro C, Arick MA, Peterson DG, Leal-Bertioli SC (2022) Variation in cytonuclear expression accommodation among allopolyploid plants. bioRxiv.
Grover CE, Gallagher JP, Szadkowski EP, Page JT, Gore MA, Udall JA, Wendel JF (2017) Nucleotide diversity in the two co-resident genomes of allopolyploid cotton. Plant systematics and evolution 303 (8):1021-1042.
Grover CE, Gallagher JP, Szadkowski EP, Yoo MJ, Flagel LE, Wendel JF (2012b) Homoeolog expression bias and expression level dominance in allopolyploids. New Phytologist 196 (4):966-971.
Grover CE, Kim H, Wing RA, Paterson AH, Wendel JF (2007) Microcolinearity and genome evolution in the AdhA region of diploid and polyploid cotton (*Gossypium*). The Plant Journal 50 (6):995-1006.
Gualberto JM, Newton KJ (2017) Plant mitochondrial genomes: dynamics and mechanisms of mutation. Annual review of plant biology 68:225-252.
Guan X, Nah G, Song Q, Udall JA, Stelly DM, Chen ZJ (2014) Transcriptome analysis of extant cotton progenitors revealed tetraploidization and identified genome-specific single nucleotide polymorphism in diploid and allotetraploid cotton. BMC research notes 7:493.
Guerra D, Crosatti C, Khoshro HH, Mastrangelo AM, Mica E, Mazzucotelli E (2015) Post-transcriptional and post-translational regulations of drought and heat response in plants: a spider's web of mechanisms. Frontiers in plant science 6:57.
Guo H, Wang X, Gundlach H, Mayer KF, Peterson DG, Scheffler BE, Chee PW, Paterson AH (2014) Extensive and biased intergenomic nonreciprocal DNA exchanges shaped a nascent polyploid genome, *Gossypium* (cotton). Genetics 197 (4):1153-1163.
Guo M, Yang S, Rupe M, Hu B, Bickel DR, Arthur L, Smith O (2008) Genome-wide allele-specific expression analysis using massively parallel signature sequencing (MPSS™) reveals cis-and trans-effects on gene expression in maize hybrid meristem tissue. Plant molecular biology 66 (5):551-563.

Gygi SP, Rochon Y, Franza BR, Aebersold R (1999) Correlation between protein and mRNA abundance in yeast. Molecular and cellular biology 19 (3):1720-1730.
Gyorfy MF, Miller ER, Conover JL, Grover CE, Wendel JF, Sloan DB, Sharbrough J (2021) Nuclear-cytoplasmic balance: whole genome duplications induce elevated organellar genome copy number. Plant Journal 108 (1):219-230.
Ha M, Kim E-D, Chen ZJ (2009) Duplicate genes increase expression diversity in closely related species and allopolyploids. Proceedings of the national Academy of sciences 106 (7):2295-2300.
Ha M, Ng DW, Li W-H, Chen ZJ (2011) Coordinated histone modifications are associated with gene expression variation within and between species. Genome research 21 (4):590-598.
Hammani K, Gobert A, Hleibieh K, Choulier L, Small I, Giegé P (2011) An *Arabidopsis* dual-localized pentatricopeptide repeat protein interacts with nuclear proteins involved in gene expression regulation. The Plant Cell 23 (2):730-740.
Havird JC, Trapp P, Miller CM, Bazos I, Sloan DB (2017) Causes and consequences of rapidly evolving mtDNA in a plant lineage. Genome biology and evolution 9 (2):323.
He G, Zhu X, Elling AA, Chen L, Wang X, Guo L, Liang M, He H, Zhang H, Chen F (2010) Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids. The Plant Cell 22 (1):17-33.
He HJ, Xie W, Liang ZX, Wu H, Bai M (2021) The expansion of mesophyll cells is coordinated with the division of chloroplasts in diploid and tetraploid *Arabidopsis thaliana*. Planta 253 (3).
Hegarty MJ, Barker GL, Wilson ID, Abbott RJ, Edwards KJ, Hiscock SJ (2006) Transcriptome shock after interspecific hybridization in *Senecio* is ameliorated by genome duplication. Current biology 16 (16):1652-1659.
Hejnák V, Hniličková H, Hnilička F, Andr J (2016) Gas exchange and *Triticum* sp. with different ploidy in relation to irradiance. Plant, Soil and Environment 62 (2):47-52.
Henry IM, Dilkes BP, Young K, Watson B, Wu H, Comai L (2005) Aneuploidy and genetic variation in the *Arabidopsis thaliana* triploid response. Genetics 170 (4):1979-1988.
Hiesey WM, Nobs MA, Björkman O (1971) Experimental studies on the nature of species. 5. Biosystematics, genetics, and physiological ecology of the Erythranthe section of *Mimulus*. na.
Hill GE (2017) The mitonuclear compatibility species concept. The Auk: Ornithological Advances 134 (2):393-409.
Hirsch J, Lefort V, Vankersschaver M, Boualem A, Lucas A, Thermes C, d'Aubenton-Carafa Y, Crespi M (2006) Characterization of 43 non-protein-coding mRNA genes in *Arabidopsis*, including the MIR162a-derived transcripts. Plant physiology 140 (4):1192-1204.
Hollister JD, Gaut BS (2009) Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome research 19 (8):1419-1428.
Hollister JD, Smith LM, Guo Y-L, Ott F, Weigel D, Gaut BS (2011) Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. Proceedings of the national Academy of sciences 108 (6):2322-2327.
Hovav R, Udall JA, Chaudhary B, Rapp R, Flagel L, Wendel JF (2008) Partitioned expression of duplicated genes during development and evolution of a single cell in a polyploid plant. Proceedings of the national Academy of sciences 105 (16):6191-6195.
Hoya A, Shibaike H, Morita T, Ito M (2007) Germination characteristics of native Japanese dandelion autopolyploids and their putative diploid parent species. Journal of Plant Research 120 (1):139-147.
Hu G, Houston NL, Pathak D, Schmidt L, Thelen JJ, Wendel JF (2011) Genomically biased accumulation of seed storage proteins in allopolyploid cotton. Genetics 189 (3):1103-1115.
Hu G, Koh J, Yoo MJ, Grupp K, Chen S, Wendel JF (2013) Proteomic profiling of developing cotton fibers from wild and domesticated *Gossypium barbadense*. New Phytologist 200 (2):570-582.

Hu G, Wendel JF (2019a) Cis–trans controls and regulatory novelty accompanying allopolyploidization. New Phytologist 221 (4):1691-1700.
Hu GJ, Wendel JF (2019b) Cis-trans controls and regulatory novelty accompanying allopolyploidization. New Phytologist 221 (4):1691-1700.
Huerta-Cepas J, Dopazo J, Huynen MA, Gabaldón T (2011) Evidence for short-time divergence and long-time conservation of tissue-specific expression after gene duplication. Briefings in bioinformatics 12 (5):442-448.
Hughes TE, Langdale JA, Kelly S (2014) The impact of widespread regulatory neofunctionalization on homeolog gene evolution following whole-genome duplication in maize. Genome research 24 (8):1348-1355.
Jackson R (1982) Polyploidy and diploidy: new perspectives on chromosome pairing and its evolutionary implications. American journal of botany 69 (9):1512-1523.
Jackson S, Chen ZJ (2010) Genomic and expression plasticity of polyploidy. Current opinion in plant biology 13 (2):153-159.
Jaskani MJ, Kwon SW, Kim DH (2005) Comparative study on vegetative, reproductive and qualitative traits of seven diploid and tetraploid watermelon lines. Euphytica 145 (3):259-268.
Jenuwein T, Allis CD (2001) Translating the histone code. Science 293 (5532):1074-1080.
Ježilová E, Nožková-Hlaváčková V, Duchoslav M (2015) Photosynthetic characteristics of three ploidy levels of *Allium oleraceum L.*(Amaryllidaceae) differing in ecological amplitude. Plant Species Biology 30 (3):212-224.
Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS (2011a) Ancestral polyploidy in seed plants and angiosperms. Nature 473 (7345):97-100.
Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS (2011b) Ancestral polyploidy in seed plants and angiosperms. Nature 473 (7345):97.
Jin Y, Liu F, Huang W, Sun Q, Huang X (2019) Identification of reliable reference genes for qRT-PCR in the ephemeral plant *Arabidopsis pumila* based on full-length transcriptome data. Scientific reports 9 (1):8408.
Johnston IG, Williams BP (2016) Evolutionary inference across eukaryotes identifies specific pressures favoring mitochondrial gene retention. Cell systems 2 (2):101-111.
Joly-Lopez Z, Bureau TE (2014) Diversity and evolution of transposable elements in *Arabidopsis*. Chromosome Research 22 (2):203-216.
Jorgensen G (1928) The experimental formation of heteroploid plants in the genus *Solanum*. Journal of Genetics 19 (2):133.
Joshi A, Van de Peer Y, Michoel T (2011) Structural and functional organization of RNA regulons in the post-transcriptional regulatory network of yeast. Nucleic acids research 39 (21):9108-9117.
Kagale S, Koh C, Nixon J, Bollina V, Clarke WE, Tuteja R, Spillane C, Robinson SJ, Links MG, Clarke C (2014) The emerging biofuel crop *Camelina sativa* retains a highly undifferentiated hexaploid genome structure. Nature communications 5:3706.
Kalyna M, Simpson CG, Syed NH, Lewandowska D, Marquez Y, Kusenda B, Marshall J, Fuller J, Cardle L, McNicol J (2011) Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in *Arabidopsis*. Nucleic acids research 40 (6):2454-2469.
Karniely S, Pines O (2005) Single translation—dual destination: mechanisms of dual protein targeting in eukaryotes. EMBO reports 6 (5):420-425.
Kashkush K, Feldman M, Levy AA (2002a) Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. Genetics 160 (4):1651-1659.
Kashkush K, Feldman M, Levy AA (2002b) Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. Nature genetics 33 (1):102.
Katoh K, Standley DM (2016) A simple method to control over-alignment in the MAFFT multiple sequence alignment program. Bioinformatics 32 (13):1933-1942.

Kazan K (2003) Alternative splicing and proteome diversity in plants: the tip of the iceberg has just emerged. Trends in plant science 8 (10):468-471.
Keene JD (2007) RNA regulons: coordination of post-transcriptional events. Nature Reviews Genetics 8 (7):533.
Kelemen O, Convertini P, Zhang Z, Wen Y, Shen M, Falaleeva M, Stamm S (2013) Function of alternative splicing. Gene 514 (1):1-30.
Keller M, Hu Y, Mesihovic A, Fragkostefanakis S, Schleiff E, Simm S (2017) Alternative splicing in tomato pollen in response to heat stress. DNA research 24 (2):205-217.
Keren H, Lev-Maor G, Ast G (2010) Alternative splicing and evolution: diversification, exon definition and function. Nature Reviews Genetics 11 (5):345.
Kim E, Magen A, Ast G (2006) Different levels of alternative splicing among eukaryotes. Nucleic acids research 35 (1):125-131.
Kleine T, Voigt C, Leister D (2009) Plastid signalling to the nucleus: messengers still lost in the mists? Trends in Genetics 25 (4):185-192.
Klodmann J, Senkler M, Rode C, Braun H-P (2011) Defining the protein complex proteome of plant mitochondria. Plant physiology 157 (2):587-598.
Klodmann J, Sunderhaus S, Nimtz M, Jänsch L, Braun H-P (2010) Internal architecture of mitochondrial complex I from *Arabidopsis thaliana*. The Plant Cell 22 (3):797-810.
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 (1):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 (1):97.
Köhler C, Scheid OM, Erilova A (2010) The impact of the triploid block on the origin and evolution of polyploid plants. Trends in Genetics 26 (3):142-148.
Kornblihtt AR, de la Mata M, Fededa JP, Munoz MJ, Nogues G (2004) Multiple links between transcription and splicing. Rna 10 (10):1489-1498.
Kornblihtt AR, Schor IE, Alló M, Dujardin G, Petrillo E, Muñoz MJ (2013) Alternative splicing: a pivotal step between eukaryotic transcription and translation. Nature reviews Molecular cell biology 14 (3):153.
Kraitshtein Z, Yaakov B, Khasdan V, Kashkush K (2010) Genetic and epigenetic dynamics of a retrotransposon after allopolyploidization of wheat. Genetics 186 (3):801-812.
Krause K, Herrmann U, Fuß J, Miao Y, Krupinska K (2009) Whirly proteins as communicators between plant organelles and the nucleus?
Krause K, Krupinska K (2009) Nuclear regulators with a second home in organelles. Trends in plant science 14 (4):194-199.
Krebbers E, Seurinck J, Herdies L, Cashmore AR, Timko MP (1988) Four genes in two diverged subfamilies encode the ribulose-1, 5-bisphosphate carboxylase small subunit polypeptides of *Arabidopsis thaliana*. Plant molecular biology 11 (6):745-759.
Lan M, Chen J-h, Fei C, Xu Q-w, Tong Z-k, Huang H-h, Dong R-h, Lou X-z, Lin E-p (2020) Induction and characterization of polyploids from seeds of Rhododendron *fortunei Lindl*. Journal of integrative agriculture 19 (8):2016-2026.
Lange C, Zerulla K, Breuert S, Soppa J (2011) Gene conversion results in the equalization of genome copies in the polyploid *haloarchaeon Haloferax volcanii*. Molecular microbiology 80 (3):666-677.
Lashermes P, Hueber Y, Combes M-C, Severac D, Dereeper A (2016) Inter-genomic DNA exchanges and homeologous gene silencing shaped the nascent allopolyploid coffee genome (*Coffea arabica L.*). Genes, genomes, genetics 6 (9):2937-2948.
Lavania UC, Srivastava S, Lavania S, Basu S, Misra NK, Mukai Y (2012) Autopolyploidy differentially influences body size in plants, but facilitates enhanced accumulation of secondary metabolites, causing increased cytosine methylation. The Plant Journal 71 (4):539-549.
Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature Reviews Genetics 11 (3):204.

Le Comber S, Ainouche M, Kovarik A, Leitch A (2010) Making a functional diploid: from polysomic to disomic inheritance. New Phytologist 186 (1):113-122.
Lee H-S, Chen ZJ (2001) Protein-coding genes are epigenetically regulated in *Arabidopsis* polyploids. Proceedings of the national Academy of sciences 98 (12):6753-6758.
Lee KH, Kim DH, Lee SW, Kim ZH, Hwang I (2002) In vivo import experiments in protoplasts reveal the importance of the overall context but not specific amino acid residues of the transit peptide during import into chloroplasts. Molecules and cells 14 (3).
Leister D (2012) Retrograde signaling in plants: from simple to complex scenarios. Frontiers in plant science 3:135.
Leitch A, Leitch I (2008) Genomic plasticity and the diversity of polyploid plants. Science 320 (5875):481-483.
Leitch IJ, Leitch AR (2013) Genome size diversity and evolution in land plants. In: Plant genome diversity, vol 2. Springer, pp 307-322.
Leon P, Arroyo A, Mackenzie S (1998) Nuclear control of plastid and mitochondrial development in higher plants. Annual review of plant biology 49 (1):453-480.
Levin DA (2002) The role of chromosomal change in plant evolution. Oxford University Press.
Levin DA (2004) The ecological transition in speciation. New Phytologist 161 (1):91-96.
Lewis BP, Green RE, Brenner SE (2003) Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. Proceedings of the national Academy of sciences 100 (1):189-192.
Lewis WH (1980) Polyploidy in species populations. In: Polyploidy. Springer, pp 103-144.
Li A, Liu D, Wu J, Zhao X, Hao M, Geng S, Yan J, Jiang X, Zhang L, Wu J (2014) mRNA and small RNA transcriptomes reveal insights into dynamic homoeolog regulation of allopolyploid heterosis in nascent hexaploid wheat. The Plant Cell 26 (5):1878-1900.
Li B, Carey M, Workman JL (2007) The role of chromatin during transcription. Cell 128 (4):707-719.
Li W-D, Biswas DK, Xu H, Xu C-Q, Wang X-Z, Liu J-K, Jiang G-M (2009) Photosynthetic responses to chromosome doubling in relation to leaf anatomy in *Lonicera japonica* subjected to water stress. Functional plant biology 36 (9):783-792.
Li Z, Defoort J, Tasdighian S, Maere S, Van de Peer Y, De Smet R (2016) Gene Duplicability of Core Genes Is Highly Consistent across All Angiosperms. Plant Cell 28 (2):326-344.
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 (10):e3353.
Ling Z, Zhou W, Baldwin IT, Xu S (2015) Insect herbivory elicits genome-wide alternative splicing responses in *Nicotiana attenuata*. The Plant Journal 84 (1):228-243.
Liu B, Brubaker C, Mergeai G, Cronn R, Wendel J (2001) Polyploid formation in cotton is not accompanied by rapid genomic changes. Genome 44 (3):321-330.
Liu C, Lu F, Cui X, Cao X (2010) Histone methylation in higher plants. Annual review of plant biology 61:395-420.
Liu Z, Qin J, Tian X, Xu S, Wang Y, Li H, Wang X, Peng H, Yao Y, Hu Z (2018) Global profiling of alternative splicing landscape responsive to drought, heat and their combination in wheat (*Triticum aestivum L.*). Plant biotechnology journal 16 (3):714-726.
Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology 15 (12):550.
Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature 389 (6648):251.
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 (1):336-348.
Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290 (5494):1151-1155.
Lynch M, Marinov GK (2015) The bioenergetic costs of a gene. Proceedings of the national Academy of sciences 112 (51):15690-15695.

Madlung A (2013a) Polyploidy and its effect on evolutionary success: old questions revisited with new tools. Heredity 110 (2):99.
Madlung A (2013b) Polyploidy and its effect on evolutionary success: old questions revisited with new tools. Heredity 110 (2):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 (2):733-746.
Madlung A, Wendel J (2013) Genetic and epigenetic aspects of polyploid evolution in plants. Cytogenetic and Genome Research 140 (2-4):270-285.
Madon M, Clyde M, Hashim H, Mohd Y, Mat H, Saratha S (2005) Polyploidy induction of oil palm through colchicine and oryzalin treatments. Journal of Oil Palm Research 17 (December):110-123.
Maherali H, Walden AE, Husband BC (2009) Genome duplication and the evolution of physiological responses to water stress. New Phytologist 184 (3):721-731.
Mandáková T, Joly S, Krzywinski M, Mummenhoff K, Lysak MA (2010) Fast diploidization in close mesopolyploid relatives of *Arabidopsis*. The Plant Cell 22 (7):2277-2290.
Maosong L, Chunyan W, Jiqing S, Yonggang C, Xiufen W, Yongfeng W (2008) Evolutional trends of leaf stomatal and photosynthetic characteristics in wheat evolutions. Acta Ecologica Sinica 28 (11):5385-5391.
Marquardt S, Raitskin O, Wu Z, Liu F, Sun Q, Dean C (2014) Functional consequences of splicing of the antisense transcript COOLAIR on FLC transcription. Molecular cell 54 (1):156-165.
Martin DP, Miya J, Reeser JW, Roychowdhury S (2016) Targeted RNA Sequencing Assay to Characterize Gene Expression and Genomic Alterations. Journal of visualized experiments (114):e54090.
Masterson J (1994) Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. Science 264 (5157):421-424.
Matyasek R, Lim K, Kovarik A, Leitch A (2003) Ribosomal DNA evolution and gene conversion in *Nicotiana rustica*. Heredity 91 (3):268.
Mavrodiev EV, Albach DC, Speranza P (2008a) A new polyploid species of the genus *Tragopogon* (Asteraceae, Cichorieae) from Russia. Novon: A Journal for Botanical Nomenclature 18 (2):229-232.
Mavrodiev EV, Chester M, Sußrez-Santiago VN, Visger CJ, Rodriguez R, Susanna A, Baldini RM, Soltis PS, Soltis DE (2015) Multiple origins and chromosomal novelty in the allotetraploid *Tragopogon castellanus* (Asteraceae). New Phytologist 206 (3):1172-1183.
Mavrodiev EV, Soltis PS, Soltis DE (2008b) Putative parentage of six Old World polyploids in *Tragopogon L.* (Asteraceae: Scorzonerinae) based on ITS, ETS, and plastid sequence data. Taxon 57 (4):1215-1212E.
Mayrose I, Zhan SH, Rothfels CJ, Magnuson-Ford K, Barker MS, Rieseberg LH, Otto SP (2011) Recently formed polyploid plants diversify at lower rates. Science 333 (6047):1257-1257.
McClintock B (1983) The significance of responses of the genome to challenge.
McGrath CL, Gout J-F, Johri P, Doak TG, Lynch M (2014) Differential retention and divergent resolution of duplicate genes following whole-genome duplication. Genome research 24 (10):1665-1675.
Meng Y, Shao C, Ma X, Wang H (2013) Introns targeted by plant microRNAs: a possible novel mechanism of gene regulation. Rice 6 (1):8.
Mercer TR, Clark MB, Crawford J, Brunck ME, Gerhardt DJ, Taft RJ, Nielsen LK, Dinger ME, Mattick JS (2014) Targeted sequencing for gene discovery and quantification using RNA CaptureSeq. Nature protocols 9 (5):989.
Mercer TR, Gerhardt DJ, Dinger ME, Crawford J, Trapnell C, Jeddeloh JA, Mattick JS, Rinn JL (2012) Targeted RNA sequencing reveals the deep complexity of the human transcriptome. Nature Biotechnology 30 (1):99-104.
Meyer EH (2012) Proteomic investigations of complex I composition: how to define a subunit? Frontiers in plant science 3:106.

Millar AH, Eubel H, Jänsch L, Kruft V, Heazlewood JL, Braun H-P (2004) Mitochondrial cytochrome c oxidase and succinate dehydrogenase complexes contain plant specific subunits. Plant molecular biology 56 (1):77-90.

Millen RS, Olmstead RG, Adams KL, Palmer JD, Lao NT, Heggie L, Kavanagh TA, Hibberd JM, Gray JC, Morden CW (2001) Many parallel losses of infA from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. The Plant Cell 13 (3):645-658.

Miller M, Zhang C, Chen ZJ (2012) Ploidy and hybridity effects on growth vigor and gene expression in *Arabidopsis thaliana* hybrids and their parents. G3 (Bethesda) 2 (4):505-513.

Mochida K, Yamazaki Y, Ogihara Y (2004) Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags. Molecular Genetics and Genomics 270 (5):371-377.

Mok D, Peloquin S (1975) The inheritance of three mechanisms of diplandroid (2n pollen) formation in diploid potatoes. Heredity 35 (3):295.

Morgner N, Zickermann V, Kerscher S, Wittig I, Abdrakhmanova A, Barth H-D, Brutschy B, Brandt U (2008) Subunit mass fingerprinting of mitochondrial complex I. Biochimica et Biophysica Acta (BBA)-Bioenergetics 1777 (10):1384-1391.

Mower JP, Bonen L (2009) Ribosomal protein L10 is encoded in the mitochondrial genome of many land plants and green algae. BMC Evolutionary Biology 9 (1):1-10.

Murat F, Zhang R, Guizard S, Flores R, Armero A, Pont C, Steinbach D, Quesneville H, Cooke R, Salse J (2013) Shared subgenome dominance following polyploidization explains grass genome evolutionary plasticity from a seven protochromosome ancestor with 16K protogenes. Genome biology and evolution 6 (1):12-33.

Murti RH, Kim HY, Yeoung YR (2012) Morphological and Anatomical Characters of Ploidy Mutants of Strawberry. International Journal of Agriculture and Biology 14 (2):204-210.

Myers W, Hill HD (1940) Studies of chromosomal association and behavior and occurrence of aneuploidy in autotetraploid grass species, orchard grass, tall oat grass, and crested wheatgrass. Botanical Gazette 102 (2):236-255.

Nalam VJ, Vales MI, Watson CJ, Kianian SF, Riera-Lizarazu O (2006) Map-based analysis of genes affecting the brittle rachis character in tetraploid wheat (*Triticum turgidum L.*). Theoretical and Applied Genetics 112 (2):373-381.

Newton nC, Pellew n (1929) *Primula kewensis* and its derivatives. Journal of Genetics 20 (3):405-467.

Newton W, Darlington C (1929) Meiosis in polyploids-Part I. Triploid and pentaploid tulips. Journal of Genetics 21 (01).

Ni Z, Kim E-D, Ha M, Lackey E, Liu J, Zhang Y, Sun Q, Chen ZJ (2009) Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. Nature 457 (7227):327.

Nie L, Han Z, Lu L, Yao Y, Sun Q, Ni Z (2008) Genomic and genic sequence variation in synthetic hexaploid wheat (AABBDD) as compared to their parental species. Progress in Natural Science 18 (5):533-538.

Niu L, Tao Y-B, Chen M-S, Fu Q, Dong Y, He H, Xu Z-F (2016) Identification and characterization of tetraploid and octoploid *Jatropha curcas* induced by colchicine. Caryologia 69 (1):58-66.

Novak SJ, Soltis DE, Soltis PS (1991) Ownbey's *Tragopogons*: 40 years later. American journal of botany 78 (11):1586-1600.

Nyikó T, Sonkoly B, Mérai Z, Benkovics AH, Silhavy D (2009) Plant upstream ORFs can trigger nonsense-mediated mRNA decay in a size-dependent manner. Plant molecular biology 71 (4-5):367.

Ohta M, Sugita M, Sugiura M (1995) Three types of nuclear genes encoding chloroplast RNA-binding proteins (cp29, cp31 and cp33) are present in *Arabidopsis thaliana*: presence of cp31 in chloroplasts and its homologue in nuclei/cytoplasms. Plant molecular biology 27 (3):529-539.

Oliver KR, McComb JA, Greene WK (2013) Transposable elements: powerful contributors to angiosperm evolution and diversity. Genome biology and evolution 5 (10):1886-1901.

Osada N, Akashi H (2011) Mitochondrial–nuclear interactions and accelerated compensatory evolution: evidence from the primate cytochrome c oxidase complex. Molecular Biology and Evolution 29 (1):337-346.
Osada N, Akashi H (2012a) Mitochondrial-Nuclear Interactions and Accelerated Compensatory Evolution: Evidence from the Primate Cytochrome c Oxidase Complex. Molecular Biology and Evolution 29 (1):337-346.
Osada N, Akashi H (2012b) Mitochondrial–nuclear interactions and accelerated compensatory evolution: evidence from the primate cytochrome c oxidase complex. Molecular Biology and Evolution 29 (1):337-346.
Osborn TC, Butrulle DV, Sharpe AG, Pickering KJ, Parkin IA, Parker JS, Lydiate DJ (2003a) Detection and effects of a homeologous reciprocal transposition in *Brassica napus*. Genetics 165 (3):1569-1577.
Osborn TC, Pires JC, Birchler JA, Auger DL, Chen ZJ, Lee H-S, Comai L, Madlung A, Doerge R, Colot V (2003b) Understanding mechanisms of novel gene expression in polyploids. Trends in Genetics 19 (3):141-147.
Ottens F, Gehring NH (2016) Physiological and pathophysiological role of nonsense-mediated mRNA decay. Pflügers Archiv-European Journal of Physiology 468 (6):1013-1028.
Otto SP (2007) The evolutionary consequences of polyploidy. Cell 131 (3):452-462.
Otto SP, Whitton J (2000) Polyploid incidence and evolution. Annual review of genetics 34 (1):401-437.
Ownbey M (1950) Natural hybridization and amphiploidy in the genus *Tragopogon*. American journal of botany:487-499.
Ownbey M, McCollum GD (1953) Cytoplasmic inheritance and reciprocal amphiploidy in *Tragopogon*. American journal of botany:788-796.
Ownbey M, McCollum GD (1954) The chromosomes of *Tragopogon*. Rhodora 56 (661):7-21.
Page JT, Liechty ZS, Alexander RH, Clemons K, Hulse-Kemp AM, Ashrafi H, Van Deynze A, Stelly DM, Udall JA (2016) DNA sequence evolution and rare homoeologous conversion in tetraploid cotton. PLoS genetics 12 (5):e1006012.
Parisi G, Perales M, Fornasari M, Colaneri A, Schain N, Casati D, Zimmermann S, Brennicke A, Araya A, Ferry J (2004) Gamma carbonic anhydrases in plant mitochondria. Plant molecular biology 55 (2):193-207.
Parisod C, Alix K, Just J, Petit M, Sarilar V, Mhiri C, Ainouche M, Chalhoub B, Grandbastien MA (2010a) Impact of transposable elements on the organization and function of allopolyploid genomes. New Phytologist 186 (1):37-45.
Parisod C, Holderegger R, Brochmann C (2010b) Evolutionary consequences of autopolyploidy. New Phytologist 186 (1):5-17.
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 Phytologist 184 (4):1003-1015.
Parkin IA, Gulden SM, Sharpe AG, Lukens L, Trick M, Osborn TC, Lydiate DJ (2005) Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. Genetics 171 (2):765-781.
Parkin IA, Koh C, Tang H, Robinson SJ, Kagale S, Clarke WE, Town CD, Nixon J, Krishnakumar V, Bidwell SL (2014) Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid *Brassica oleracea*. Genome biology 15 (6):R77.
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 (11):597-602.
Pécrix 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 (10):3587-3597.
Peng I, Ronin Y, Fahima T, Röder M, Li Y, Nevo E, Korol A Genomic distribution of domestication QTLs in wild emmer wheat, *Triticum dicoccoides*. In: Proceedings of the 10th international wheat genetics symposium, 2003a. pp 34-37

Peng J, Ronin Y, Fahima T, Röder MS, Li Y, Nevo E, Korol A (2003b) Domestication quantitative trait loci in *Triticum dicoccoides*, the progenitor of wheat. Proceedings of the national Academy of sciences 100 (5):2489-2494.
Perales M, Parisi G, Fornasari MS, Colaneri A, Villarreal F, Gonzalez-Schain N, Echave J, Gomez-Casati D, Braun H-P, Araya A (2004) Gamma carbonic anhydrase like complex interact with plant mitochondrial complex I. Plant molecular biology 56 (6):947-957.
Phillips JP, Tainer JA, Getzoff ED, Boulianne GL, Kirby K, Hilliker AJ (1995) Subunit-destabilizing mutations in *Drosophila* copper/zinc superoxide dismutase: neuropathology and a model of dimer dysequilibrium. Proceedings of the national Academy of sciences 92 (19):8574-8578.
Pikaard CS, Scheid OM (2014) Epigenetic regulation in plants. Cold Spring Harbor perspectives in biology 6 (12):a019315.
Pires JC, Zhao J, Schranz ME, Leon EJ, Quijada PA, Lukens LN, Osborn TC (2004) Flowering time divergence and genomic rearrangements in resynthesized *Brassica* polyploids (Brassicaceae). Biological Journal of the Linnean Society 82 (4):675-688.
Pogson BJ, Woo NS, Förster B, Small ID (2008) Plastid signalling to the nucleus and beyond. Trends in plant science 13 (11):602-609.
Pophaly SD, Tellier A (2015) Population level purifying selection and gene expression shape subgenome evolution in maize. Molecular Biology and Evolution 32 (12):3226-3235.
Postel Z, Touzet P (2020) Cytonuclear genetic incompatibilities in plant speciation. Plants 9 (4):487.
Prince VE, Pickett FB (2002) Splitting pairs: the diverging fates of duplicated genes. Nature Reviews Genetics 3 (11):827.
Pumphrey M, Bai J, Laudencia-Chingcuanco D, Anderson O, Gill BS (2009) Nonadditive expression of homoeologous genes is established upon polyploidization in hexaploid wheat. Genetics 181 (3):1147-1157.
Qi B, Huang W, Zhu B, Zhong X, Guo J, Zhao N, Xu C, Zhang H, Pang J, Han F (2012) Global transgenerational gene expression dynamics in two newly synthesized allohexaploid wheat (*Triticum aestivum*) lines. BMC biology 10 (1):3.
Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, Zhang S, Paterson AH (2019) Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. Genome biology 20 (1):38.
Quinlan AR (2014) BEDTools: the Swiss-army tool for genome feature analysis. Current protocols in bioinformatics 47 (1):11.12. 11-11.12. 34.
Radhamony RN, Theg SM (2006) Evidence for an ER to Golgi to chloroplast protein transport pathway. Trends in cell biology 16 (8):385-387.
Rambani A, Page JT, Udall JA (2014) Polyploidy and the petal transcriptome of *Gossypium*. BMC Plant Biology 14 (1):3.
Ramsey J, Ramsey TS (2014) Ecological studies of polyploidy in the 100 years following its discovery. Philosophical Transactions of the Royal Society B: Biological Sciences 369 (1648):20130352.
Ramsey J, Schemske DW (1998) Pathways, mechanisms, and rates of polyploid formation in flowering plants. Annual review of ecology and systematics 29 (1):467-501.
Ramsey J, Schemske DW (2002) Neopolyploidy in flowering plants. Annual review of ecology and systematics 33 (1):589-639.
Rand DM, Haney RA, Fry AJ (2004) Cytonuclear coevolution: the genomics of cooperation. Trends in ecology & evolution 19 (12):645-653.
Rapp RA, Udall JA, Wendel JF (2009a) Genomic expression dominance in allopolyploids. BMC biology 7 (1):18.
Rapp RA, Udall JA, Wendel JF (2009b) Genomic expression dominance in allopolyploids. Bmc Biology 7.
Reddy AS (2007) Alternative splicing of pre-messenger RNAs in plants in the genomic era. Annual review of plant biology 58:267-294.
Reddy AS, Marquez Y, Kalyna M, Barta A (2013) Complexity of the alternative splicing landscape in plants. The Plant Cell 25 (10):3657-3683.

Reddy AS, Rogers MF, Richardson DN, Hamilton M, Ben-Hur A (2012) Deciphering the plant splicing code: experimental and computational approaches for predicting alternative splicing and splicing regulatory elements. Frontiers in plant science 3:18.
Ren Y, Li Y, Jiang Y, Wu B, Miao Y (2017) Phosphorylation of WHIRLY1 by CIPK14 shifts its localization and dual functions in *Arabidopsis*. Molecular plant 10 (5):749-763.
Renny-Byfield S, Gallagher JP, Grover CE, Szadkowski E, Page JT, Udall JA, Wang X, Paterson AH, Wendel JF (2014) Ancient gene duplicates in *Gossypium* (cotton) exhibit near-complete expression divergence. Genome biology and evolution 6 (3):559-571.
Renny-Byfield S, Gong L, Gallagher JP, Wendel JF (2015) Persistence of subgenomes in paleopolyploid cotton after 60 my of evolution. Molecular Biology and Evolution 32 (4):1063-1071.
Renny-Byfield S, Rodgers-Melnick E, Ross-Ibarra J (2017) Gene fractionation and function in the ancient subgenomes of maize. Molecular Biology and Evolution 34 (8):1825-1832.
Rhoads DM, Subbaiah CC (2007) Mitochondrial retrograde regulation in plants. Mitochondrion 7 (3):177-194.
Richards EJ (1997) DNA methylation and plant development. Trends in Genetics 13 (8):319-323.
Riddle NC, Birchler JA (2003) Effects of reunited diverged regulatory hierarchies in allopolyploids and species hybrids. Trends in Genetics 19 (11):597-600.
Rockenbach K, Havird JC, Monroe JG, Triant DA, Taylor DR, Sloan DB (2016) Positive selection in rapidly evolving plastid–Nuclear enzyme complexes. Genetics 204 (4):1507-1522.
Rodrigues FA, Soares JDR, dos Santos RR, Pasqual M, e Silva S (2011) Colchicine and amiprophos-methyl (APM) in polyploidy induction in banana plant. African Journal of Biotechnology 10 (62):13476-13481.
Romero-Aranda R, Bondada B, Syvertsen J, Grosser J (1997) Leaf characteristics and net gas exchange of diploid and autotetraploid citrus. Annals of Botany 79 (2):153-160.
Roose M, Gottlieb L (1976) Genetic and biochemical consequences of polyploidy in *Tragopogon*. Evolution 30 (4):818-830.
Rosloski SM, Singh A, Jali SS, Balasubramanian S, Weigel D, Grbic V (2013) Functional analysis of splice variant expression of MADS AFFECTING FLOWERING 2 of *Arabidopsis thaliana*. Plant molecular biology 81 (1-2):57-69.
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 (1):143-153.
Roux F, Mary-Huard T, Barillot E, Wenes E, Botran L, Durand S, Villoutreix R, Martin-Magniette M-L, Camilleri C, Budar F (2016) Cytonuclear interactions affect adaptive traits of the annual plant *Arabidopsis thaliana* in the field. Proceedings of the national Academy of sciences 113 (13):3687-3692.
Ruiz M, Quiñones A, Martínez-Alcántara B, Aleza P, Morillon R, Navarro L, Primo-Millo E, Martínez-Cuenca M-R (2016) Tetraploidy enhances boron-excess tolerance in *carrizo citrange* (*Citrus sinensis L. Osb*.× *Poncirus trifoliata L. Raf.*). Frontiers in plant science 7:701.
Sablok G, Gupta P, Baek J-M, Vazquez F, Min XJ (2011) Genome-wide survey of alternative splicing in the grass *Brachypodium distachyon*: a emerging model biosystem for plant functional genomics. Biotechnology letters 33 (3):629-636.
Sablok G, Powell B, Braessler J, Yu F, Min XJ (2017) Comparative landscape of alternative splicing in fruit plants. Current Plant Biology 9:29-36.
Salinas-Giegé T, Giegé R, Giegé P (2015) tRNA biology in mitochondria. International journal of molecular sciences 16 (3):4518-4559.
Salman-Minkov A, Sabath N, Mayrose I (2016a) Whole-genome duplication as a key factor in crop domestication. Nature plants 2 (8):1-4.
Salman-Minkov A, Sabath N, Mayrose I (2016b) Whole-genome duplication as a key factor in crop domestication. Nature plants 2 (8):16115.
Salmon A, Ainouche ML (2010) Polyploidy and DNA methylation: new tools available. Molecular ecology 19 (2):213-215.

Salmon A, Ainouche ML, Wendel JF (2005) Genetic and epigenetic consequences of recent hybridization and polyploidy in *Spartina* (Poaceae). Molecular ecology 14 (4):1163-1175.
Salmon A, Flagel L, Ying B, Udall JA, Wendel JF (2010) Homoeologous nonreciprocal recombination in polyploid cotton. New Phytologist 186 (1):123-134.
Saminathan T, Nimmakayala P, Manohar S, Malkaram S, Almeida A, Cantrell R, Tomason Y, Abburi L, Rahman MA, Vajja VG (2014) Differential gene expression and alternative splicing between diploid and tetraploid watermelon. Journal of experimental botany 66 (5):1369-1385.
Sánchez-Fernández Ro, Davies TE, Coleman JO, Rea PA (2001) The *Arabidopsis thaliana* ABC protein superfamily, a complete inventory. Journal of Biological Chemistry 276 (32):30231-30244.
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 (1):161-171.
Sattler MC, Carvalho CR, Clarindo WR (2016) The polyploidy and its key role in plant breeding. Planta 243 (2):281-296.
Sazanov LA (2015) A giant molecular proton pump: structure and mechanism of respiratory complex I. Nature reviews Molecular cell biology 16 (6):375-388.
Scheid OM, Afsar K, Paszkowski J (2003) Formation of stable epialleles and their paramutation-like interaction in tetraploid *Arabidopsis thaliana*. Nature genetics 34 (4):450.
Schnable JC, Springer NM, Freeling M (2011) Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proceedings of the national Academy of sciences 108 (10):4069-4074.
Schnable PS, Wise RP (1998) The molecular basis of cytoplasmic male sterility and fertility restoration. Trends in plant science 3 (5):175-180.
Sears ER (1954) The aneuploids of common wheat. University of Missouri, College of Agriculture, Agricultural Experiment Station.
Seguí-Simarro JM, Coronado MJ, Staehelin LA (2008) The mitochondrial cycle of *Arabidopsis* shoot apical meristem and leaf primordium meristematic cells is defined by a perinuclear tentaculate/cage-like mitochondrion. Plant physiology 148 (3):1380-1393.
Sehrish T, Symonds VV, Soltis DE, Soltis PS, Tate JA (2014) Gene silencing via DNA methylation in naturally occurring *Tragopogon miscellus* (Asteraceae) allopolyploids. BMC genomics 15 (1):701.
Sehrish T, Symonds VV, Soltis DE, Soltis PS, Tate JA (2015) Cytonuclear coordination is not immediate upon allopolyploid formation in *Tragopogon miscellus* (Asteraceae) allopolyploids. PloS one 10 (12):e0144339.
Selvi F, Vivona L (2022) Polyploidy in *Odontarrhena bertolonii* (Brassicaceae) in relation to seed germination performance and plant phenotype, with taxonomic implications. Plant biosystems 156 (4):959-968.
Semeniuk P, Arisumi T (1968) Colchicine-induced tetraploid and cytochimeral roses. Botanical Gazette 129 (3):190-193.
Senerchia N, Felber F, Parisod C (2015) Genome reorganization in F1 hybrids uncovers the role of retrotransposons in reproductive isolation. Proceedings of the Royal Society B: Biological Sciences 282 (1804):20142874.
Senkler J, Senkler M, Eubel H, Hildebrandt T, Lengwenus C, Schertl P, Schwarzländer M, Wagner S, Wittig I, Braun HP (2017) The mitochondrial complexome of *Arabidopsis thaliana*. The Plant Journal 89 (6):1079-1092.
Seo PJ, Park M-J, Lim M-H, Kim S-G, Lee M, Baldwin IT, Park C-M (2012) A self-regulatory circuit of CIRCADIAN CLOCK-ASSOCIATED1 underlies the circadian clock regulation of temperature responses in *Arabidopsis*. The Plant Cell 24 (6):2427-2442.
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 (9):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. The Plant Cell 13 (8):1749-1759.
Shamandi N, Zytnicki M, Charbonnel C, Elvira-Matelot E, Bochnakian A, Comella P, Mallory AC, Lepere G, Saez-Vasquez J, Vaucheret H (2015) Plants encode a general siRNA suppressor that is induced and suppressed by viruses. PLoS biology 13 (12):e1002326.
Shan S, Boatwright JL, Liu X, Chanderbali AS, Fu C, Soltis PS, Soltis DE (2020a) Transcriptome Dynamics of the Inflorescence in Reciprocally Formed Allopolyploid Tragopogon miscellus (Asteraceae). Frontiers in Genetics 11.
Shan S, Boatwright JL, Liu X, Chanderbali AS, Fu C, Soltis PS, Soltis DE (2020b) Transcriptome Dynamics of the Inflorescence in Reciprocally Formed Allopolyploid Tragopogon miscellus (Asteraceae). Front Genet 11:888.
Shan S, Boatwright JL, Liu X, Chanderbali AS, Fu C, Soltis PS, Soltis DE (2020c) Transcriptome dynamics of the inflorescence in reciprocally formed allopolyploid *Tragopogon miscellus* (Asteraceae). Frontiers in genetics 11:888.
Sharbrough J, Conover JL, Gyorfy MF, Grover CE, Miller ER, Wendel JF, Sloan DB (2022) Global patterns of subgenome evolution in organelle-targeted genes of six allotetraploid angiosperms. Molecular Biology and Evolution 39 (4).
Sharbrough J, Conover JL, Tate JA, Wendel JF, Sloan DB (2017) Cytonuclear responses to genome doubling. American journal of botany 104 (9).
Shen Y, Zhou Z, Wang Z, Li W, Fang C, Wu M, Ma Y, Liu T, Kong L-A, Peng D-L (2014) Global dissection of alternative splicing in paleopolyploid soybean. The Plant Cell 26 (3):996-1008.
Shi X, Ng DW, Zhang C, Comai L, Ye W, Chen ZJ (2012a) Cis-and trans-regulatory divergence between progenitor species determines gene-expression novelty in *Arabidopsis* allopolyploids. Nature communications 3:950.
Shi X, Zhang C, Ko DK, Chen ZJ (2015) Genome-wide dosage-dependent and-independent regulation contributes to gene expression and evolutionary novelty in plant polyploids. Molecular Biology and Evolution 32 (9):2351-2366.
Shi XL, Ng DWK, Zhang CQ, Comai L, Ye WX, Chen ZJ (2012b) Cis- and trans-regulatory divergence between progenitor species determines gene-expression novelty in Arabidopsis allopolyploids. Nature communications 3.
Shin KH, Yang SH, Lee JY, Lim CW, Lee SC, Brown JW, Kim SH (2015) Alternative splicing of mini-exons in the *Arabidopsis* leaf rust receptor-like kinase LRK10 genes affects subcellular localisation. Plant cell reports 34 (3):495-505.
Shrestha B, Gilbert LE, Ruhlman TA, Jansen RK (2020) Rampant nuclear transfer and substitutions of plastid genes in *Passiflora*. Genome biology and evolution 12 (8):1313-1329.
Sigel EM, Der JP, Windham MD, Pryer KM (2019) Expression level dominance and homeolog expression bias in recurrent origins of the allopolyploid fern *Polypodium hesperium*. American Fern Journal 109 (3):224-247.
Silva-Filho MC (2003) One ticket for multiple destinations: dual targeting of proteins to distinct subcellular locations. Current opinion in plant biology 6 (6):589-595.
Simioni C, Schifino-Wittmann MT, Dall'Agnol M (2006) Sexual polyploidization in red clover. Scientia Agricola 63 (1):26-31.
Singh B, Gupta RS (2006) Mitochondrial import of human and yeast fumarase in live mammalian cells: retrograde translocation of the yeast enzyme is mainly caused by its poor targeting sequence. Biochemical and biophysical research communications 346 (3):911-918.
Singh B, Soltys BJ, Wu Z-C, Patel HV, Freeman KB, Gupta RS (1997) Cloning and some novel characteristics of mitochondrial Hsp70 from Chinese hamster cells. Experimental cell research 234 (2):205-216.
Skalińska M (1947) Polyploidy in *Valeriana officinalis Linn.* in relation to its ecology and distribution. Botanical journal of the Linnean Society 53 (350):159-186.
Skelly MJ, Frungillo L, Spoel SH (2016) Transcriptional regulation by complex interplay between post-translational modifications. Current opinion in plant biology 33:126-132.

Skippington E, Barkman TJ, Rice DW, Palmer JD (2015) Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all nad genes. Proceedings of the national Academy of sciences 112 (27):E3515-E3524.
Slater GSC, Birney E (2005) Automated generation of heuristics for biological sequence comparison. BMC bioinformatics 6 (1):1-11.
Sloan DB (2015) Using plants to elucidate the mechanisms of cytonuclear co-evolution. New Phytologist 205 (3):1040-1046.
Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, Taylor DR (2012) Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. PLoS biology 10 (1):e1001241.
Sloan DB, Triant DA, Wu M, Taylor DR (2014) Cytonuclear interactions and relaxed selection accelerate sequence evolution in organelle ribosomes. Molecular Biology and Evolution 31 (3):673-682.
Sloan DB, Warren JM, Williams AM, Wu Z, Abdel-Ghany SE, Chicco AJ, Havird JC (2018a) Cytonuclear integration and co-evolution. Nature reviews genetics 19 (10):635-648.
Sloan DB, Warren JM, Williams AM, Wu Z, Abdel-Ghany SE, Chicco AJ, Havird JC (2018b) Cytonuclear integration and co-evolution. Nature Reviews Genetics 19 (10):635.
Snodgrass S, Jareczek J, Wendel J (2017) An examination of nucleotypic effects in diploid and polyploid cotton. AoB Plants 9 (1):plw082.
Soltis D, Soltis P, Rieseberg LH (1993) Molecular data and the dynamic nature of polyploidy. Critical reviews in plant sciences 12 (3):243-273.
Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, de Pamphilis CW, Wall PK, Soltis PS (2009) Polyploidy and angiosperm diversification. American journal of botany 96 (1):336-348.
Soltis DE, Buggs RJ, Barbazuk WB, Chamala S, Chester M, Gallagher JP, Schnable PS, Soltis PS (2012) The early stages of polyploidy: rapid and repeated evolution in *Tragopogon*. In: Polyploidy and genome evolution. Springer, pp 271-292.
Soltis DE, Soltis PS (1989) Allopolyploid speciation in *Tragopogon*: insights from chloroplast DNA. American journal of botany 76 (8):1119-1124.
Soltis DE, Soltis PS (1995) The dynamic nature of polyploid genomes. Proceedings of the national Academy of sciences 92 (18):8089.
Soltis DE, Soltis PS (1999) Polyploidy: recurrent formation and genome evolution. Trends in ecology & evolution 14 (9):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 (4):485-501.
Soltis DE, Soltis PS, Schemske DW, Hancock JF, Thompson JN, Husband BC, Judd WS (2007) Autopolyploidy in angiosperms: have we grossly underestimated the number of species? Taxon 56 (1):13-30.
Soltis DE, Visger CJ, Marchant DB, Soltis PS (2016) Polyploidy: pitfalls and paths to a paradigm. American journal of botany 103 (7):1146-1166.
Soltis DE, Visger CJ, Soltis PS (2014a) The polyploidy revolution then… and now: Stebbins revisited. American journal of botany 101 (7):1057-1078.
Soltis PS, Liu X, Marchant DB, Visger CJ, Soltis DE (2014b) Polyploidy and novelty: Gottlieb's legacy. Philosophical Transactions of the Royal Society B: Biological Sciences 369 (1648):20130351.
Soltis PS, Marchant DB, Van de Peer Y, Soltis DE (2015) Polyploidy and genome evolution in plants. Current opinion in genetics & development 35:119-125.
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 (10):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 (10):1329-1341.

Soltis PS, Soltis DE (1991) Multiple origins of the allotetraploid *Tragopogon mirus* (Compositae): rDNA evidence. Systematic Botany: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 97 (13):7051-7057.
Soltis PS, Soltis DE (2016) Ancient WGD events as drivers of key innovations in angiosperms. Current opinion in plant biology 30:159-165.
Soltys BJ, Gupta RS (1996) Immunoelectron microscopic localization of the 60-kDa heat shock chaperonin protein (Hsp60) in mammalian cells. Experimental cell research 222 (1):16-27.
Soltys BJ, Gupta RS (1999) Mitochondrial-matrix proteins at unexpected locations: are they exported? Trends in biochemical sciences 24 (5):174-177.
Song K, Lu P, Tang K, Osborn TC (1995) Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. Proceedings of the national Academy of sciences 92 (17):7719-7723.
Song Q, Chen ZJ (2015) Epigenetic and developmental regulation in plant polyploids. Current opinion in plant biology 24:101-109.
Spoelhof JP, Chester M, Rodriguez R, Geraci B, Heo K, Mavrodiev E, Soltis PS, Soltis DE (2017) Karyotypic variation and pollen stainability in resynthesized allopolyploids *Tragopogon miscellus* and *T. mirus*. American journal of botany 104 (10):1484-1492.
Springer NM, Lisch D, Li Q (2016) Creating order from chaos: epigenome dynamics in plants with complex genomes. The Plant Cell 28 (2):314-325.
Sriswasdi S, Takashima M, Manabe R-i, Ohkuma M, Sugita T, Iwasaki W (2016) Global deceleration of gene evolution following recent genome hybridizations in fungi. Genome research 26 (8):1081-1090.
Srivalli B, Khanna-Chopra R (2004) Ribulose-1, 5-bisphosphate carboxylase/oxygenase content and degradation in diploid, tetraploid, and hexaploid wheat species during monocarpic senescence. Photosynthetica 42 (3):393-398.
Staiger D, Brown JW (2013) Alternative splicing at the intersection of biological timing, development, and stress responses. The Plant Cell 25 (10):3640-3656.
Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30 (9):1312-1313.
Stamm S, Ben-Ari S, Rafalska I, Tang Y, Zhang Z, Toiber D, Thanaraj T, Soreq H (2005) Function of alternative splicing. Gene 344:1-20.
Stebbins (1950) Variation and evolution in plants. Variation and evolution in plants.
Stebbins GL (1971a) Chromosomal evolution in higher plants. Chromosomal evolution in higher plants.
Stebbins GLJCeihp (1971b) Chromosomal evolution in higher plants.
Stebbins Jr GL (1947) Types of polyploids: their classification and significance. In: Advances in genetics, vol 1. Elsevier, pp 403-429.
Steige KA, Slotte T (2016) Genomic legacies of the progenitors and the evolutionary consequences of allopolyploidy. Current opinion in plant biology 30:88-93.
Stevens A, Nicotra A, Godfree R, Guja LK (2020) Polyploidy affects the seed, dormancy and seedling characteristics of a perennial grass, conferring an advantage in stressful climates. Plant Biology 22 (3):500-513.
Sturgill D, Malone JH, Sun X, Smith HE, Rabinow L, Samson M-L, Oliver B (2013) Design of RNA splicing analysis null models for post hoc filtering of Drosophila head RNA-Seq data with the splicing analysis kit (Spanki). BMC bioinformatics 14 (1):320.
Sugiyama S-I (2005) Polyploidy and cellular mechanisms changing leaf size: comparison of diploid and autotetraploid populations in two species of *Lolium*. Annals of Botany 96 (5):931-938.
Sultan M, Schulz MH, Richard H, Magen A, Klingenhoff A, Scherf M, Seifert M, Borodina T, Soldatov A, Parkhomchuk D (2008) A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science 321 (5891):956-960.

Sun H, Wu S, Zhang G, Jiao C, Guo S, Ren Y, Zhang J, Zhang H, Gong G, Jia Z (2017) Karyotype stability and unbiased fractionation in the paleo-allotetraploid *Cucurbita* genomes. Molecular plant 10 (10):1293-1306.

Sun Q, Sun H, Bell RL, Li L, Zhou G, Xin L, Wei Z (2015) Field performance of vegetative form traits of neopolyploids produced by in vitro colchicine treatment in *Pyrus communis*. Scientia Horticulturae 193:182-187.

Sunderland PA, West CE, Waterworth WM, Bray CM (2006) An evolutionarily conserved translation initiation mechanism regulates nuclear or mitochondrial targeting of DNA ligase 1 in *Arabidopsis thaliana*. The Plant Journal 47 (3):356-367.

Syed NH, Kalyna M, Marquez Y, Barta A, Brown JW (2012) Alternative splicing in plants–coming of age. Trends in plant science 17 (10):616-623.

Symonds VV, Soltis PS, Soltis DE (2010) Dynamics of Polyploid Formation in Tragopogon (Asteraceae): Recurrent Formation, Gene Flow, and Population Structure. Evolution 64 (7):1984-2003.

Szarzynska B, Sobkowiak L, Pant BD, Balazadeh S, Scheible W-R, Mueller-Roeber B, Jarmolowski A, Szweykowska-Kulinska Z (2009) Gene structures and processing of *Arabidopsis thaliana* HYL1-dependent pri-miRNAs. Nucleic acids research 37 (9):3083-3093.

Takahagi K, Inoue K, Shimizu M, Uehara-Yamaguchi Y, Onda Y, Mochida K (2018a) Homoeolog-specific activation of genes for heat acclimation in the allopolyploid grass Brachypodium hybridum. Gigascience 7 (4).

Takahagi K, Inoue K, Shimizu M, Uehara-Yamaguchi Y, Onda Y, Mochida K (2018b) Homoeolog-specific activation of genes for heat acclimation in the allopolyploid grass *Brachypodium hybridum*. GigaScience 7 (4):giy020.

Talbert PB, Henikoff S (2006) Spreading of silent chromatin: inaction at a distance. Nature Reviews Genetics 7 (10):793.

Tamayo-Ordóñez M, Espinosa-Barrera L, Tamayo-Ordóñez Y, Ayil-Gutiérrez B, Sánchez-Teyer L (2016) Advances and perspectives in the generation of polyploid plant species. Euphytica 209 (1):1-22.

Tang H, Woodhouse MR, Cheng F, Schnable JC, Pedersen BS, Conant G, Wang X, Freeling M, Pires JC (2012) Altered patterns of fractionation and exon deletions in *Brassica rapa* support a two-step model of paleohexaploidy. Genetics 190 (4):1563-1574.

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 (1):1-10.

Tate JA, Joshi P, Soltis KA, Soltis PS, Soltis DE (2009b) On the road to diploidization? Homoeolog loss in independently formed populations of the allopolyploid *Tragopogon miscellus* (Asteraceae). BMC Plant Biology 9 (1):80.

Tate JA, Ni Z, Scheen A-C, Koh J, Gilbert CA, Lefkowitz D, Chen ZJ, Soltis PS, Soltis DE (2006a) Evolution and expression of homeologous loci in *Tragopogon miscellus* (Asteraceae), a recent and reciprocally formed allopolyploid. Genetics 173 (3):1599-1611.

Tate JA, Ni ZF, Scheen AC, 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 (3):1599-1611.

Tate JA, Soltis DE, Soltis PS (2005a) The evolution of the genome.

Tate JA, Soltis DE, Soltis PS (2005b) Polyploidy in plants. In: The evolution of the genome. Elsevier, pp 371-426.

Tate JA, Symonds VV, Doust AN, Buggs RJ, Mavrodiev E, Majure LC, Soltis PS, Soltis DE (2009c) Synthetic polyploids of *Tragopogon miscellus* and *T. mirus* (Asteraceae): 60 years after Ownbey's discovery. American journal of botany 96 (5):979-988.

Tayalé A, Parisod C (2013) Natural pathways to polyploidy in plants and consequences for genome reorganization. Cytogenetic and Genome Research 140 (2-4):79-96.

Taylor WC (1989) Regulatory interactions between nuclear and plastid genomes. Annual review of plant biology 40 (1):211-233.
Thomas BC, Pedersen B, Freeling M (2006) Following tetraploidy in an *Arabidopsis* ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. Genome research 16 (7):934-946.
Timmis JN, Ayliffe MA, Huang CY, Martin W (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nature Reviews Genetics 5 (2):123-135.
Tjalsma H, Bolhuis A, Jongbloed JD, Bron S, van Dijl JM (2000) Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome. Microbiology and molecular biology reviews 64 (3):515-547.
Udall JA, Quijada PA, Osborn TC (2005) Detection of chromosomal rearrangements derived from homeologous recombination in four mapping populations of *Brassica napus L*. Genetics 169 (2):967-979.
Uversky VN (2015) Functional roles of transiently and intrinsically disordered regions within proteins. The FEBS journal 282 (7):1182-1189.
Uversky VN (2016) Dancing protein clouds: the strange biology and chaotic physics of intrinsically disordered proteins. Journal of Biological Chemistry 291 (13):6681-6688.
Vallejo-Marín M, Buggs RJ, Cooley AM, Puzey JR (2015) Speciation by genome duplication: Repeated origins and genomic composition of the recently formed allopolyploid species *Mimulus peregrinus*. Evolution 69 (6):1487-1500.
Van de Peer Y, Mizrachi E, Marchal K (2017a) The evolutionary significance of polyploidy. Nature Reviews Genetics 18 (7):411.
Van de Peer Y, Mizrachi E, Marchal K (2017b) The evolutionary significance of polyploidy. Nature Reviews Genetics 18 (7):411-424.
van der Sluis EO, Bauerschmitt H, Becker T, Mielke T, Frauenfeld J, Berninghausen O, Neupert W, Herrmann JM, Beckmann R (2015) Parallel Structural Evolution of Mitochondrial Ribosomes and OXPHOS Complexes. Genome Biology and Evolution 7 (5):1235-1251.
Van Laere K, França 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 (4):1149-1156.
Verhoeven KJ, Jansen JJ, van Dijk PJ, Biere A (2010) Stress-induced DNA methylation changes and their heritability in asexual dandelions. New Phytologist 185 (4):1108-1118.
Vincent D, Lapierre C, Pollet B, Cornic G, Negroni L, Zivy M (2005) Water deficits affect caffeate O-methyltransferase, lignification, and related enzymes in maize leaves. A proteomic investigation. Plant physiology 137 (3):949-960.
Von Well E, Fossey A (1998) A comparative investigation of seed germination, metabolism and seedling growth between two polyploid *Triticum* species. Euphytica 101 (1):83-89.
Vyas P, Bisht MS, Miyazawa S-I, Yano S, Noguchi K, Terashima I, Funayama-Noguchi S (2007) Effects of polyploidy on photosynthetic properties and anatomy in leaves of *Phlox drummondii*. Functional plant biology 34 (8):673-682.
Wagner A (2005) Distributed robustness versus redundancy as causes of mutational robustness. Bioessays 27 (2):176-188.
Wang B-B, Brendel V (2006) Genomewide comparative analysis of alternative splicing in plants. Proceedings of the national Academy of sciences 103 (18):7175-7180.
Wang B, Tseng E, Regulski M, Clark TA, Hon T, Jiao Y, Lu Z, Olson A, Stein JC, Ware D (2016) Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. Nature communications 7:11708.
Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB (2008) Alternative isoform regulation in human tissue transcriptomes. Nature 456 (7221):470.
Wang J, Tian L, Lee H-S, Chen ZJ (2006a) Nonadditive regulation of FRI and FLC loci mediates flowering-time variation in *Arabidopsis* allopolyploids. Genetics 173 (2):965-974.
Wang J, Tian L, Lee H-S, Wei NE, Jiang H, Watson B, Madlung A, Osborn TC, Doerge R, Comai L (2006b) Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. Genetics 172 (1):507-517.

Wang J, Tian L, Madlung A, Lee H-S, Chen M, Lee JJ, Watson B, Kagochi T, Comai L, Chen ZJ (2004) Stochastic and epigenetic changes of gene expression in *Arabidopsis* polyploids. Genetics 167 (4):1961-1973.
Wang JL, Tian L, Lee HS, Chen ZJ (2006c) Nonadditive regulation of FRI and FLC loci mediates flowering-time variation in Arabidopsis allopolyploids. Genetics 173 (2):965-974.
Wang JL, Tian L, Lee HS, Wei NE, Jiang HM, Watson B, Madlung A, Osborn TC, Doerge RW, Comai L, Chen ZJ (2006d) Genomewide nonadditive gene regulation in Arabidopsis allotetraploids. Genetics 172 (1):507-517.
Wang M, Wang P, Liang F, Ye Z, Li J, Shen C, Pei L, Wang F, Hu J, Tu L (2018) A global survey of alternative splicing in allopolyploid cotton: landscape, complexity and regulation. New Phytologist 217 (1):163-178.
Wang R, Liu H, Liu Z, Zou J, Meng J, Wang J (2019) Genome-wide analysis of alternative splicing divergences between *Brassica* hexaploid and its parents. Planta:1-26.
Wang S, Chen Y (2018) Phylogenomic analysis demonstrates a pattern of rare and long-lasting concerted evolution in prokaryotes. Communications biology 1 (1):12.
Wang S, Chen Y (2019) Fine-tuning the expression of duplicate genes by translational regulation in *Arabidopsis* and maize. Frontiers in plant science 10:534.
Wang X-Y, Paterson AH (2011) Gene conversion in angiosperm genomes with an emphasis on genes duplicated by polyploidization. Genes 2 (1):1-20.
Wang X, Dong Q, Li X, Yuliang A, Yu Y, Li N, Liu B, Gong L (2017a) Cytonuclear variation of rubisco in synthesized rice hybrids and allotetraploids. The Plant Genome 10 (3):plantgenome2017.2005.0041.
Wang X, Dong Q, Li X, Yuliang A, Yu Y, Li N, Liu B, Gong L (2017b) Cytonuclear variation of rubisco in synthesized rice hybrids and allotetraploids. The plant genome 10 (3).
Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun J-H, Bancroft I, Cheng F (2011) The genome of the mesopolyploid crop species *Brassica rapa*. Nature genetics 43 (10):1035.
Wang Y, Wang X, Lee TH, Mansoor S, Paterson AH (2013) 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 (1):274-283.
Wang Z, Burge CB (2008) Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. Rna 14 (5):802-813.
Warner DA, Edwards GE (1993) Effects of polyploidy on photosynthesis. Photosynthesis Research 35 (2):135-147.
Weeden NF (1981) Genetic and biochemical implications of the endosymbiotic origin of the chloroplast. Journal of Molecular Evolution 17 (3):133-139.
Weiss H, Maluszynska J (2001) Chromosomal rearrangement in autotetraploid plants of *Arabidopsis thaliana*. Hereditas 133 (3):255-261.
Wendel J, Doyle J (2005) Polyploidy and evolution in plants. In: Plant diversity and evolution: genotypic and phenotypic variation in higher plants. CABI publishing Wallingford UK, pp 97-117.
Wendel JF (2000) Genome evolution in polyploids. Plant molecular biology 42 (1):225-249.
Wendel JF (2015) The wondrous cycles of polyploidy in plants. American journal of botany 102 (11):1753-1756.
Wendel JF, Cronn RC (2003a) Polyploidy and the evolutionary history of cotton. Advances in agronomy 78:139.
Wendel JF, Cronn RC (2003b) 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: Polyploidy and genome evolution. Springer, pp 181-207.
Wendel JF, Lisch D, Hu G, Mason AS (2018a) The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation. Current opinion in genetics & development 49:1-7.

Wendel JF, Lisch D, Hu GJ, Mason AS (2018b) The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation. Current opinion in genetics & development 49:1-7.
Whittle CA, Krochko JE (2009) Transcript profiling provides evidence of functional divergence and expression networks among ribosomal protein gene paralogs in *Brassica napus*. The Plant Cell 21 (8):2203-2219.
Wicke S, Schneeweiss GM, Depamphilis CW, Müller KF, Quandt D (2011) The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant molecular biology 76:273-297.
Wolfe KH (2001) Yesterday's polyploids and the mystery of diploidization. Nature Reviews Genetics 2 (5):333.
Wolfe KH, Li W-H, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proceedings of the national Academy of sciences 84 (24):9054-9058.
Wolfe TM, Balao F, Trucchi E, Bachmann G, Gu W, Baar J, Hedrén M, Weckwerth W, Leitch AR, Paun O (2021) Recurrent allopolyploidization events diversify eco-physiological traits in marsh orchids. bioRxiv.
Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH (2009) The frequency of polyploid speciation in vascular plants. Proceedings of the national Academy of sciences 106 (33):13875-13879.
Woodhouse MR, Cheng F, Pires JC, Lisch D, Freeling M, Wang X (2014) Origin, inheritance, and gene regulatory consequences of genome dominance in polyploids. Proceedings of the national Academy of sciences 111 (14):5283-5288.
Wu J-H, Ferguson AR, Murray BG, Jia Y, Datson PM, Zhang J (2012) Induced polyploidy dramatically increases the size and alters the shape of fruit in *Actinidia chinensis*. Annals of Botany 109 (1):169-179.
Wu S, Han B, Jiao Y (2020) Genetic contribution of paleopolyploidy to adaptive evolution in angiosperms. Molecular plant 13 (1):59-71.
Xiao-Ming Z, Junrui W, Li F, Sha L, Hongbo P, Lan Q, Jing L, Yan S, Weihua Q, Lifang Z (2017) Inferring the evolutionary mechanism of the chloroplast genome size by comparing whole-chloroplast genome sequences in seed plants. Scientific reports 7 (1):1555.
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 (19):7908-7913.
Xu C, Bai Y, Lin X, Zhao N, Hu L, Gong Z, Wendel JF, Liu B (2014) Genome-wide disruption of gene expression in allopolyploids but not hybrids of rice subspecies. Molecular Biology and Evolution 31 (5):1066-1076.
Xu F, Xu S, Wiermer M, Zhang Y, Li X (2012) The cyclin L homolog MOS12 and the MOS4-associated complex are required for the proper splicing of plant resistance genes. The Plant Journal 70 (6):916-928.
Xu L, Carrie C, Law SR, Murcha MW, Whelan J (2013) Acquisition, conservation, and loss of dual-targeted proteins in land plants. Plant physiology 161 (2):644-662.
Xue H, Zhang B, Tian J-R, Chen M-M, Zhang Y-Y, Zhang Z-H, Ma Y (2017) Comparison of the morphology, growth and development of diploid and autotetraploid 'Hanfu'apple trees. Scientia Horticulturae 225:277-285.
Yan K, Liu P, Wu C-A, Yang G-D, Xu R, Guo Q-H, Huang J-G, Zheng C-C (2012) Stress-induced alternative splicing provides a mechanism for the regulation of microRNA processing in *Arabidopsis thaliana*. Molecular cell 48 (4):521-531.
Yang X, Zhang H, Li L (2012) Alternative mRNA processing increases the complexity of microRNA-based gene regulation in *Arabidopsis*. The Plant Journal 70 (3):421-431.
Yao H, Kato A, Mooney B, Birchler JA (2011) Phenotypic and gene expression analyses of a ploidy series of maize inbred Oh43. Plant molecular biology 75 (3):237-251.
Yeo GW, Van Nostrand EL, Liang TY (2007) Discovery and analysis of evolutionarily conserved intronic splicing regulatory elements. PLoS genetics 3 (5):e85.

Yoo M-J, Liu X, Pires JC, Soltis PS, Soltis DE (2014a) Nonadditive gene expression in polyploids. Annual review of genetics 48:485-517.
Yoo M, Szadkowski E, Wendel J (2013a) Homoeolog expression bias and expression level dominance in allopolyploid cotton. Heredity 110 (2):171.
Yoo MJ, Liu XX, Pires JC, Soltis PS, Soltis DE (2014b) Nonadditive Gene Expression in Polyploids. Annual Review of Genetics, Vol 48 48:485-517.
Yoo MJ, Szadkowski E, Wendel JF (2013b) Homoeolog expression bias and expression level dominance in allopolyploid cotton. Heredity 110 (2):171-180.
Yoshimura K, Yabuta Y, Ishikawa T, Shigeoka S (2002) Identification of a cis element for tissue-specific alternative splicing of chloroplast ascorbate peroxidase pre-mRNA in higher plants. Journal of Biological Chemistry 277 (43):40623-40632.
Yu X, Zhai Y, Wang P, Cheng C, Li J, Lou Q, Chen J (2021) Morphological, anatomical and photosynthetic consequences of artificial allopolyploidization in *Cucumis*. Euphytica 217 (1):5.
Zhai Y, Yu X, Zhu Z, Wang P, Meng Y, Zhao Q, Li J, Chen J (2019a) Nuclear-Cytoplasmic Coevolution Analysis of RuBisCO in Synthesized Cucumis Allopolyploid. Genes (Basel) 10 (11).
Zhai Y, Yu X, Zhu Z, Wang P, Meng Y, Zhao Q, Li J, Chen J (2019b) Nuclear–cytoplasmic coevolution analysis of RuBisCO in synthesized *Cucumis* allopolyploid. Genes 10 (11):869.
Zhang D, Yang J-F, Gao B, Liu T-Y, Hao G-F, Yang G-F, Fu L-J, Chen M-X, Zhang J (2019) Identification, evolution and alternative splicing profile analysis of the splicing factor 30 (SPF30) in plant species. Planta 249 (6):1997-2014.
Zhang F, Xue H, Lu X, Zhang B, Wang F, Ma Y, Zhang Z (2015a) Autotetraploidization enhances drought stress tolerance in two apple cultivars. Trees 29 (6):1773-1780.
Zhang G, Guo G, Hu X, Zhang Y, Li Q, Li R, Zhuang R, Lu Z, He Z, Fang X (2010a) Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. Genome research 20 (5):646-654.
Zhang J, Liu Y, Xia E-H, Yao Q-Y, Liu X-D, Gao L-Z (2015b) Autotetraploid rice methylome analysis reveals methylation variation of transposable elements and their effects on gene expression. Proceedings of the national Academy of sciences 112 (50):E7022-E7029.
Zhang J, Ruhlman TA, Sabir JS, Blazier JC, Weng M-L, Park S, Jansen RK (2016) Coevolution between nuclear-encoded DNA replication, recombination, and repair genes and plastid genome complexity. Genome biology and evolution 8 (3):622-634.
Zhang L-Q, Liu D-C, Zheng Y-L, Yan Z-H, Dai S-F, Li Y-F, Jiang Q, Ye Y-Q, Yen Y (2010b) Frequent occurrence of unreduced gametes in *Triticum turgidum–Aegilops tauschii* hybrids. Euphytica 172 (2):285-294.
Zhang W, Fan X, Gao Y, Liu L, Sun L, Su Q, Han J, Zhang N, Cui F, Ji J (2017) Chromatin modification contributes to the expression divergence of three TaGS2 homoeologs in hexaploid wheat. Scientific reports 7:44677.
Zhao J, Udall JA, Quijada PA, Grau CR, Meng J, Osborn TC (2006) Quantitative trait loci for resistance to *Sclerotinia sclerotiorum* and its association with a homeologous non-reciprocal transposition in *Brassica napus L*. Theoretical and Applied Genetics 112 (3):509-516.
Zhao J, Wang X, Deng B, Lou P, Wu J, Sun R, Xu Z, Vromans J, Koornneef M, Bonnema G (2005) Genetic relationships within *Brassica rapa* as inferred from AFLP fingerprints. Theoretical and Applied Genetics 110 (7):1301-1314.
Zhao N, Zhu B, Li M, Wang L, Xu L, Zhang H, Zheng S, Qi B, Han F, Liu B (2011) Extensive and heritable epigenetic remodeling and genetic stability accompany allohexaploidization of wheat. Genetics 188 (3):499-510.
Zhou R, Moshgabadi N, Adams KL (2011) Extensive changes to alternative splicing patterns following allopolyploidy in natural and resynthesized polyploids. Proceedings of the national Academy of sciences 108 (38):16122-16127.
Zielinski M-L, Scheid OM (2012) Meiosis in polyploid plants. In: Polyploidy and genome evolution. Springer, pp 33-55.

Zou C, Lehti-Shiu MD, Thomashow M, Shiu SH (2009) Evolution of stress-regulated gene expression in duplicate genes of *Arabidopsis thaliana*. PLoS genetics 5 (7).

Appendices

Chapter 2 Supplementary Data

Supplementary Table 2.1: List of candidate nuclear genes involved in cytonuclear interactions in *Arabidopsis thaliana*.

TAIR Gene Name	Enzyme Complex	Gene Locus
NADH-ubiquinone oxidoreductase B8 subunit	OXPHOS Complex I	At5g47890
LYR family of Fe/S cluster biogenesis protein	OXPHOS Complex I	At3g12260
Furry	OXPHOS Complex I	At5g08060
NAD(P)-binding Rossmann-fold superfamily protein	OXPHOS Complex I	At2g20360
Mitochondrial import inner membrane translocase subunit Tim17/Tim22/Tim23 family protein	OXPHOS Complex I	At2g42210
NADH:ubiquinone oxidoreductase 17.2kDa subunit	OXPHOS Complex I	At3g03100
AT2G31490 protein	OXPHOS Complex I	At2g31490
NADH-ubiquinone oxidoreductase B18 subunit	OXPHOS Complex I	At2g02050
NADH dehydrogenase	OXPHOS Complex I	At5g47570
LYR family of Fe/S cluster biogenesis protein	OXPHOS Complex I	At4g34700
Excitatory amino acid transporter	OXPHOS Complex I	At4g20150
NADH-ubiquinone dehydrogenase mitochondrial	OXPHOS Complex I	At5g37510
NADH-ubiquinone oxidoreductase-related	OXPHOS Complex I	At5g67590
NADH-ubiquinone oxidoreductase-related	OXPHOS Complex I	At3g03070
NADH-ubiquinone oxidoreductase 20 kDa subunit mitochondrial	OXPHOS Complex I	At5g11770
51 kDa subunit of complex I	OXPHOS Complex I	At5g08530
NADH-ubiquinone oxidoreductase 24 kDa subunit	OXPHOS Complex I	At4g02580
At1g67350	OXPHOS Complex I	At1g67350
copper ion binding	OXPHOS Complex I	At2g27730
Putative uncharacterized protein	OXPHOS Complex I	At1g67785
At4g16450	OXPHOS Complex I	At4g16450
Transmembrane protein	OXPHOS Complex I	At4g00585

TAIR Gene Name	Enzyme Complex	Gene Locus
L-galactono-1 4-lactone dehydrogenase	OXPHOS Complex I	At3g47930
At1g68680	OXPHOS Complex I	At1g68680
At5g14105	OXPHOS Complex I	At5g14105
NADH dehydrogenase	OXPHOS Complex I	At3g08610
NADH-ubiquinone oxidoreductase-related	OXPHOS Complex I	At5g52840
GRIM-19 protein	OXPHOS Complex I	At2g33220 and At1g04630[1]
NADH dehydrogenase (ubiquinone)s	OXPHOS Complex I	At2g02510 and At1g14450[1]
Complex I subunit NDUFS6	OXPHOS Complex I	At1g49140 and At3g18410[1]
ESSS subunit of NADH:ubiquinone oxidoreductase (Complex I) protein	OXPHOS Complex I	At2g42310 and At3g57785[1]
NADH-ubiquinone oxidoreductase-related	OXPHOS Complex I	At3g62790 and At2g47690[1]
Alpha-helical ferredoxin	OXPHOS Complex I	At1g79010 and At1g16700[1]
gamma carbonic anhydrase 1	OXPHOS Complex I	At1g19580
gamma carbonic anhydrase 2	OXPHOS Complex I	At1g47260
gamma carbonic anhydrase like 1	OXPHOS Complex I	At5g63510
gamma carbonic anhydrase-like 2	OXPHOS Complex I	At3g48680
At2g46540/F11C10.23	OXPHOS Complex I	At2g46540
Cox19-like CHCH family protein	OXPHOS Complex I	At3g06310
Cox19-like CHCH family protein	OXPHOS Complex I	At5g18800
NADH dehydrogenase	OXPHOS Complex I	At1g76200
gamma carbonic anhydrase 3	OXPHOS Complex I	At5g66510
Insulinase (Peptidase family M16) protein	OXPHOS Complex III	At3g02090
Insulinase (Peptidase family M16) protein and mitochondrial processing peptidase alpha subunit	OXPHOS Complex III	At1g51980 and At3g16480[1]
ubiquinol-cytochrome C reductase UQCRX/QCR9-like family protein	OXPHOS Complex III	At3g52730
Cytochrome C1 family	OXPHOS Complex III	At5g40810 and At3g27240[1]
Ubiquinol-cytochrome C reductase iron-sulfur subunit	OXPHOS Complex III	At5g13430 and At5g13440[1]
Cytochrome bd ubiquinol oxidase 14kDa subunit	OXPHOS Complex III	At4g32470 and At5g25450[1]

TAIR Gene Name	Enzyme Complex	Gene Locus
Cytochrome b-c1 complex subunit 8 protein	OXPHOS Complex III	At3g10860 and At5g05370[1]
Putative uncharacterized protein At2g40765	OXPHOS Complex III	At2g40765
Ubiquinol-cytochrome C reductase hinge protein	OXPHOS Complex III	At2g01090
Rubredoxin-like superfamily protein	OXPHOS Complex IV	At3g15640 and At1g80230[1]
cytochrome-c oxidases;electron carriers	OXPHOS Complex IV	At4g21105
TonB-dependent heme receptor A	OXPHOS Complex IV	At1g72020
cytochrome C oxidase 6B	OXPHOS Complex IV	At1g22450
Protein of unknown function (DUF1640)	OXPHOS Complex IV	At2g16460
Protein of unknown function (DUF1138)	OXPHOS Complex IV	At4g00860 and At1g011701
Cytochrome c oxidase subunit Vc family protein, unknown protein	OXPHOS Complex IV	At2g47380, At3g62400 and At5g61310[1]
cytochrome c oxidase-related	OXPHOS Complex IV	At4g37830
Hypoxia-responsive family protein	OXPHOS Complex IV	At5g27760
F-box/LRR protein	OXPHOS Complex IV	At3g43410
ATP synthase epsilon chain mitochondrial	OXPHOS Complex V	At1g51650
ATPase F1 complex delta/epsilon subunit	OXPHOS Complex V	At5g47030
delta subunit of Mt ATP synthase	OXPHOS Complex V	At5g13450
ATP synthase D chain mitochondrial	OXPHOS Complex V	At3g52300
copper ion binding;cobalt ion binding;zinc ion binding	OXPHOS Complex V	At2g21870
ATP synthase alpha/beta family protein	OXPHOS Complex V	At5g08670, At5g08680 and At5g08690[1]
AT4g30010/F6G3_40	OXPHOS Complex V	At4g30010
Mitochondrial ATP synthase subunit G protein	OXPHOS Complex V	At4g29480, At2g19680 and At4g26210[1]
AT3g46430/F18L15_150	OXPHOS Complex V	At3g46430 and At5g59613[1]
gamma subunit of Mt ATP synthase	OXPHOS Complex V	At2g33040
photosystem I light harvesting complex gene 1	Photosystem I	At3g54890
photosystem I light harvesting complex gene 2	Photosystem I	At3g61470
photosystem I light harvesting complex gene 3	Photosystem I	At1g61520

TAIR Gene Name	Enzyme Complex	Gene Locus
light-harvesting chlorophyll-protein complex I subunit A4	Photosystem I	At3g47470
photosystem I light harvesting complex gene 6	Photosystem I	At1g19150
Photosystem I reaction centre subunit IV / PsaE protein and photosystem I subunit E-2	Photosystem I	At4g28750 and At2g20260[1]
photosystem I subunit F	Photosystem I	At1g31330
photosystem I subunit G	Photosystem I	At1g55670
photosystem I subunit K	Photosystem I	At1g30380
photosystem I subunit l	Photosystem I	At4g12800
photosystem I reaction center subunit PSI-N chloroplast putative / PSI-N putative (PSAN)	Photosystem I	At5g64040
photosystem I P subunit	Photosystem I	At2g46820
photosystem I subunit D-1 and photosystem I subunit D-2	Photosystem I	At4g02770 and At1g03130[1]
photosystem I subunit H-1 and photosystem I subunit H-2	Photosystem I	At3g16140 and At1g52230[1]
photosystem I subunit O	Photosystem I	At1g08380
photosystem I light harvesting complex gene 5	Photosystem I	At1g45474
photosystem II subunit P-1 and photosystem II subunit P-2	Photosystem II	At1g06680 and At2g30790[1]
photosystem II subunit R	Photosystem II	At1g79040
Chlorophyll A-B binding family protein	Photosystem II	At1g44575
photosystem II reaction center W	Photosystem II	At2g30570
photosystem II subunit X	Photosystem II	At2g06520
photosystem II BY	Photosystem II	At1g67740
light-harvesting chlorophyll B-binding protein 3	Photosystem II	At5g54270
light harvesting complex of photosystem II 5	Photosystem II	At4g10340
light harvesting complex photosystem II subunit 6	Photosystem II	At1g15820
PS II oxygen-evolving complex 1 and photosystem II subunit O-2	Photosystem II	At5g66570 and At3g50820[1]
photosystem II subunit QA and photosystem II subunit Q-2	Photosystem II	At4g21280 and At4g05180[1]
chlorophyll A/B-binding protein 2, chlorophyll A/B binding protein 3, chlorophyll A/B binding protein 1, light-harvesting chlorophyll-protein complex II subunit B1 and photosystem II light harvesting complex gene B1B2	Photosystem II	At1g29920, At1g29910,At1g29930, At2g34430 and At2g34420[1]

TAIR Gene Name	Enzyme Complex	Gene Locus
light harvesting complex photosystem II	Photosystem II	At5g01530, At3g08940 and At2g40100[1]
photosystem II subunit T	Photosystem II	At3g21055
photosystem II light harvesting complex gene 2.1	Photosystem II	At2g05100
sigma factor A	RNA polymerase PEP	At1g64860
RNApolymerase sigma-subunit C	RNA polymerase PEP	At3g53920
sigma factor E	RNA polymerase PEP	At5g24120
RNApolymerase sigma-subunit F	RNA polymerase PEP	At2g36990
RNApolymerase sigma subunit 2	RNA polymerase PEP	At1g08540
sigma factor 4	RNA polymerase PEP	At5g13730
ATPase F1 complex gamma subunit protein	ATP Synthase	At4g04640
ATP synthase delta-subunit gene	ATP Synthase	At4g09650
ATPase F0 complex subunit B/B' bacterial/chloroplast	ATP Synthase	At4g32260
photosynthetic electron transfer C	Cytochrome b6f	At4g03280
cytochrome b6f complex subunit (petM) putative	Cytochrome b6f	At2g26500
CLP protease proteolytic subunit 3	Caseinolytic Proteases	At1g66670
nuclear encoded CLP protease 5	Caseinolytic Proteases	At1g02560
CLP protease proteolytic subunit 2	Caseinolytic Proteases	At1g12410
ATP-dependent caseinolytic (Clp) protease/crotonase family protein	Caseinolytic Proteases	At1g09130
CLP protease R subunit 4	Caseinolytic Proteases	At4g17040
Clp ATPase	Caseinolytic Proteases	At5g51070
Ribosomal protein L12/ ATP-dependent Clp protease adaptor protein ClpS family protein	Caseinolytic Proteases	At1g68660
CLPT1 and CLPT2	Caseinolytic Proteases	At4g25370 and At4g12060[1]
CLP protease P4	Caseinolytic Proteases	At5g45390
CLPC homologue 1 and Clp ATPase	Caseinolytic Proteases	At5g50920 and At3g48870[1]
CLP protease proteolytic subunit 6	Caseinolytic Proteases	At1g11750
CLP protease proteolytic subunit 1	Caseinolytic Proteases	At1g49970
uvrB/uvrC motif-containing protein	Caseinolytic Proteases	At2g03390

TAIR Gene Name	Enzyme Complex	Gene Locus
acetyl Co-enzyme a carboxylase carboxyltransferase alpha subunit	Acetyl-CoA Carboxylase	At2g38040
acetyl Co-enzyme a carboxylase biotin carboxylase subunit	Acetyl-CoA Carboxylase	At5g35360
chloroplastic acetylcoenzyme A carboxylase 1 and biotin carboxyl carrier protein 2	Acetyl-CoA Carboxylase	At5g16390 and At5g15530[1]
ribulose bisphosphate carboxylase small chain 1A, Ribulose bisphosphate carboxylase (small chain) family protein	Rubisco	At1g67090, At5g38410, At5g38430 and At5g38420[1]
RNA-binding (RRM/RBD/RNP motifs) family protein	Ribosomal Genes	At3g52150
plastid-specific ribosomal protein 4	Ribosomal Genes	At2g38140
plastid-specific 50S ribosomal protein 5	Ribosomal Genes	At3g56910
plastid-specific 50S ribosomal protein 6	Ribosomal Genes	At5g17870
ribosomal protein L4	Ribosomal Genes	At1g07320
Ribosomal protein L6 family	Ribosomal Genes	At1g05190
ribosomal protein L9	Ribosomal Genes	At3g44890
Ribosomal protein L10 family protein	Ribosomal Genes	At5g13510
plastid ribosomal protein l11 and Ribosomal protein L11 family protein	Ribosomal Genes	At1g32990 and At5g51610[1]
Ribosomal protein L13 family protein	Ribosomal Genes	At1g78630
ribosomal protein L15	Ribosomal Genes	At3g25920
Ribosomal protein L17 family protein	Ribosomal Genes	At3g54210
Ribosomal L18p/L5e family protein	Ribosomal Genes	At1g48350 and At3g20230[1]
Ribosomal protein L19 family protein	Ribosomal Genes	At5g47190 and At4g17560[1]
Ribosomal protein L21	Ribosomal Genes	At1g35680
Translation protein SH3-like family protein	Ribosomal Genes	At5g54600
ribosomal protein large subunit 27	Ribosomal Genes	At5g40950
Ribosomal L28 family	Ribosomal Genes	At2g33450
Ribosomal L29 family protein	Ribosomal Genes	At5g65220
Ribosomal protein L31	Ribosomal Genes	At1g75350
Ribosomal protein L34	Ribosomal Genes	At1g29070
Ribosomal protein L35	Ribosomal Genes	At2g24090

TAIR Gene Name	Enzyme Complex	Gene Locus
ribosomal protein S1	Ribosomal Genes	At5g30510
Ribosomal protein S5 family protein	Ribosomal Genes	At2g33800
Translation elongation factor EF1B/ribosomal protein S6 family protein	Ribosomal Genes	At1g64510
ribosomal protein S9	Ribosomal Genes	At1g74970
Ribosomal protein S10p/S20e family protein	Ribosomal Genes	At3g13120
Ribosomal protein S13/S18 family	Ribosomal Genes	At5g14320
ribosomal protein S17	Ribosomal Genes	At1g79850
chloroplast 30S ribosomal protein S20 putative	Ribosomal Genes	At3g15190
Ribosomal protein S21 family protein	Ribosomal Genes	At3g27160
ribosomal protein L12-A, L12-B and L12-C	Ribosomal Genes	At3g27830, At3g27840 and At3g27850[1]
30S ribosomal protein putative	Ribosomal Genes	At5g24490
Ribosomal protein PSRP-3/Ycf65	Ribosomal Genes	At1g68590
Ribosomal protein L1p/L10e family	Ribosomal Genes	At3g63490
Ribosomal protein L3 family protein	Ribosomal Genes	At2g43030
Ribosomal L5P family protein	Ribosomal Genes	At4g01310
Pentatricopeptide repeat (PPR) superfamily protein	Dual-targeted Genes	At5g60960
ssDNA-binding transcriptional regulator	Dual-targeted Genes	At1g14410
Mitochondrial transcription termination factor family protein	Dual-targeted Genes	At4g38160

Supplementary Table 2.2: List of candidate control nuclear genes that are not involved in cytonuclear interactions in *Arabidopsis thaliana*.

TAIR Gene Name	Enzyme Complex	Gene Locus
succinate dehydrogenase 2-1, 2-2, 2-3	OXPHOS Complex II	At3g27380, At5g40650 and At5g651651
succinate dehydrogenase subunit 7A mitochondrial and SDH7B	OXPHOS Complex II	At3g47833 and At5g625751
succinate dehydrogenase 1-1 and succinate dehydrogenase 1-2	OXPHOS Complex II	At5g66760 and At2g184501
succinate dehydrogenase 5	OXPHOS Complex II	At1g47420
succinate dehydrogenase subunit 6 mitochondrial	OXPHOS Complex II	At1g08480
succinate dehydrogenase 3-1 and succinate dehydrogenase 3-2	OXPHOS Complex II	At5g09600 and At4g322101
succinate dehydrogenase subunit 4	OXPHOS Complex II	At2g46505
succinate dehydrogenase subunit 8 mitochondrial	OXPHOS Complex II	At2g46390
NAD(P)H dehydrogenase B1	OXPHOS Alternative Pathway	At4g28220
NAD(P)H dehydrogenase B3	OXPHOS Alternative Pathway	At4g21490
NAD(P)H dehydrogenase B2	OXPHOS Alternative Pathway	At4g05020
NAD(P)H dehydrogenase B4	OXPHOS Alternative Pathway	At2g20800
alternative oxidase 1A, 1B, 1C and 1D	OXPHOS Alternative Pathway	At3g22370, At3g22360, At3g27620 and At1g323501
alternative oxidase 2	OXPHOS Alternative Pathway	At5g64210
alternative NAD(P)H dehydrogenase 1	OXPHOS Alternative Pathway	At1g07180
alternative NAD(P)H dehydrogenase 2	OXPHOS Alternative Pathway	At2g29990
NAD(P)H dehydrogenase C1	OXPHOS Alternative Pathway	At5g08740
NUCLEOSOME ASSEMLY PROTEIN 1/NAP1	Retrograde-signalling Genes	At4g26110
TCP family transcription factor	Retrograde-signalling Genes	At1g58100
Peptidase C78, ubiquitin fold modifier-specific peptidase 1/ 2	Retrograde-signalling Genes	At5g24680
transmembrane protein	Retrograde-signalling Genes	At1g44920
serine/threonine protein kinase 1	Retrograde-signalling Genes	At5g01820
glyceraldehyde-3-phosphate dehydrogenase C2	Housekeeping Genes	At1g13440
TIP41-like family protein	Housekeeping Genes	At4g34270
SAND family protein	Housekeeping Genes	At2g28390
Clathrin adaptor complexes medium subunit family protein	Housekeeping Genes	At5g46630
AT4g33380/F17M5_140	Housekeeping Genes	At4g33380
HAC13 protein (HAC13)	Housekeeping Genes	At1g32750

Supplementary Table 2.3: List of *Arabidopsis* nuclear genes that were missing in the *Tragopogon dubius* draft genome as well as *Lactuca sativa* genome.

TAIR Gene name	Enzyme complex	Gene locus
gamma carbonic anhydrase 3	OXPHOS Complex I	At5g66510
succinate dehydrogenase 3-1 and succinate dehydrogenase 3-2	OXPHOS Complex II	At5g09600 and At4g32210[1]
succinate dehydrogenase subunit 4	OXPHOS Complex II	At2g46505
Succinate dehydrogenase subunit 8 mitochondrial	OXPHOS Complex II	At2g46390
mitochondrial processing peptidase alpha subunit	OXPHOS Complex III	At3G16480
Putative uncharacterized protein At2g40765	OXPHOS Complex III	At2G40765
Ubiquinol-cytochrome C reductase hinge protein	OXPHOS Complex III	At2G01090
Rubredoxin-like superfamily protein	OXPHOS Complex IV	At3G15640
Cytochrome c oxidase subunit Vc family protein	OXPHOS Complex IV	At2G47380
F-box/LRR protein	OXPHOS Complex IV	At3g43410
photosystem I light harvesting complex gene 5	Photosystem I	At1g45474
Photosystem I reaction centre subunit IV / PsaE protein	Photosystem I	At4g28750
photosystem II subunit P-2	Photosystem II	At2g30790
light harvesting complex photosystem II	Photosystem II	At3g08940
light harvesting complex photosystem II	Photosystem II	At2g40100
sigma factor 4	RNA polymerase PEP	At5g13730
uvrB/uvrC motif-containing protein	Caseinolytic Proteases	At2g03390
Ribosomal protein L11 family protein	Ribosomal Genes	At5g51610
ribosomal protein L12-B	Ribosomal Genes	At3g27840
Ribosomal L18p/L5e family protein	Ribosomal Genes	At3g20230

Chapter 3 Supplementary Data

Supplementary Table 3.1: List of individuals from natural populations of diploid parent species, *Tragopogon dubius* and *T. pratensis* as well as reciprocal forms of *T. miscellus* examined.

Species	Population	Population ID	Individual ID	Replicates	Plant group
T. dubius	Moscow, Idaho, USA	106	2	2	Late-flowering
T. dubius	Oakesdale, Washington, USA	3055	12	3	Late-flowering
T. pratensis	Oakesdale, Washington, USA	3052	18	2	Late-flowering
T. pratensis	Garfield, Washington, USA	3058	8	2	Late-flowering
T. pratensis	Garfield, Washington, USA	3058	10	2	Late-flowering
T. miscellus (long-liguled)	Pullman, Washington, USA	3066	10	2	Late-flowering
T. miscellus (long-liguled)	Pullman, Washington, USA	101	5	2	Late-flowering
T. miscellus (short-liguled)	Spangle, Washington, USA	3049	1	2	Late-flowering
T. miscellus (short-liguled)	Garfield, Washington, USA	104	4	2	Late-flowering
T. miscellus (short-liguled)	Rosalia, Washington, USA	3073	7	2	Late-flowering
T. miscellus (short-liguled)	Oakesdale, Washington, USA	3053	8	3	Late-flowering
T. miscellus (short-liguled)	Oakesdale, Washington, USA	3053	9	2	Late-flowering
T. dubius	Spangle, Washington, USA	3050	1	2	Early-flowering
T. dubius	Oakesdale, Washington, USA	3055	8	3	Early-flowering
T. pratensis	Garfield, Washington, USA	3058	9	1	Early-flowering
T. miscellus (short-liguled)	Rosalia, Washington, USA	3073	6	3	Early-flowering
T. miscellus (short-liguled)	Oakesdale, Washington, USA	3053	13	3	Early-flowering

Supplementary Table 3.2: Number of duplicated nuclear genes differentially expressed between diploid parent species, *Tragopogon dubius* and *T. pratensis* at different growth stages.

Expression patterns	Early stage	Middle stage	Mature stage	Early stage-late flowering plants	Early stage-early flowering plants
No differential expression	199	197	204	198	202
Expression higher in *T. pratensis*	2	5	0	2	1
Expression higher in *T. dubius*	3	3	1	4	1
Total	204	205	205	204	204

Supplementary Table 3.3: Number of duplicated nuclear genes differentially expressed between *Tragopogon dubius* and populations of short- and long-liguled *T. miscellus* at different growth stages.

	Short-liguled *T. miscellus* (sm)														**Long-liguled *T. miscellus* (lm)**	
	104		**3049**		**3053**					**3073**					**3066+101**	
Expression patterns	**Early stage**	**Middle stage**	**Early stage**	**Middle stage**	**Early stage**	**Middle stage**	**Mature stage**	**Early stage-late flowering plants**	**Early stage-early flowering plants**	**Early stage**	**Middle stage**	**Mature stage**	**Early stage-late flowering plants**	**Early stage-early flowering plants**	**Early stage**	**Middle stage**
No differential expression	204	203	202	198	202	203	198	203	201	199	199	202	203	187	200	201
Expression higher in *T. dubius*	0	0	1	2	1	0	0	0	2	3	4	0	0	15	3	3
Expression higher in sm/lm	0	2	1	5	1	2	7	1	1	2	2	3	1	2	1	1
Total	204	205	204	205	204	205	205	204	204	204	205	205	204	204	204	205

Supplementary Table 3.4: Number of duplicated nuclear genes differentially expressed between *Tragopogon pratensis* and populations of short- and long-liguled *T. miscellus* at different growth stages.

	Short-liguled *T. miscellus* (sm)														**Long-liguled *T. miscellus* (lm)**	
	104		**3049**		**3053**					**3073**					**3066+101**	
Expression patterns	**Early stage**	**Middle stage**	**Early stage**	**Middle stage**	**Early stage**	**Middle stage**	**Mature stage**	**Early stage-late flowering plants**	**Early stage-early flowering plants**	**Early stage**	**Middle stage**	**Mature stage**	**Early stage-late flowering plants**	**Early stage-early flowering plants**	**Early stage**	**Middle stage**
No differential expression	201	203	201	203	203	202	205	201	204	203	193	205	199	200	201	198
Expression higher in *T. pratensis*	1	0	0	0	0	1	0	1	0	1	10	0	3	4	2	6
Expression higher in sm/lm	2	2	3	2	1	2	0	2	0	0	2	0	2	0	1	1
Total	204	205	204	205	204	205	205	204	204	204	205	205	204	204	204	205

Supplementary Table 3.5: Number of duplicated nuclear genes differentially expressed between long-liguled *Tragopogon miscellus* (lm) and populations of short-liguled *T. miscellus* (sm) at different growth stages.

	104		**3049**		**3053**		**3073**	
Expression patterns	**Early stage**	**Middle stage**	**Early stage**	**Middle stage**	**Early stage**	**Middle stage**	**Early stage**	**Middle stage**
No differential expression	204	201	204	201	204	203	204	204
Expression higher in sm	0	4	0	4	0	1	0	0
Expression higher in lm	0	0	0	0	0	1	0	1
Total	204	205	204	205	204	205	204	205

Supplementary Table 3.6: List of genes with anomalous expression patterns in populations (marked X) of short- and long-liguled *Tragopogon miscellus* at early-, middle- and mature-growth stages.

			Late-flowering group											Early-flowering group			
			Early stage		Early stage					Middle stage					Early vs Mature stage			
			Short-liguled *T. miscellus*						Long-liguled *T. miscellus*	Short-liguled *T. miscellus*				Long-liguled *T. miscellus*	Short-liguled *T. miscellus*			
			3053	3073	104	3049	3053	3073	3066+101	104	3049	3053	3073	3066+101	3053	3073	3053	3073
Gene ID	**Gene name**	**Enzyme complex**	**No difference in expression b/w parents & polyploid for loci differentially expressed between parent species**															
TragDub04873-RA	NADH dehydrogenase (ubiquinone)s	OXPHOS Complex I			X	X	X	X	X	X	X	X	X	X				
TragDub00693-RA	Ubiquinol-cytochrome C reductase iron-sulfur subunit	OXPHOS Complex III								X	X	X	X	X				
Lsat_20510	photosystem I subunit O	Photosystem I			X	X	X	X	X									
TragDub28661-RA	chlorophyll A/B-binding protein	Photosystem II	X															
TragDub19360-RA	chlorophyll A/B-binding protein	Photosystem II		X														
Lsat_10456	photosystem II light harvesting complex gene 2.1	Photosystem II			X	X	X	X	X			X						
TragDub22127-RA	plastid-specific ribosomal protein 4	Ribosomal Genes	X												X			
Lsat_5980	30S ribosomal protein putative	Ribosomal Genes													X			
Lsat_1652	Ribosomal protein L3 family protein	Ribosomal Genes															X	X
TragDub00311-RA	sigma factor E	RNA Polymerase PEP								X				X				
TragDub19304-RA	CBL-interacting protein kinase 14	Dual-Targeted Genes								X		X	X	X				
		No difference in expression b/w polyploid & *T. pratensis* but low expression in polyploid than *T. dubius* for loci that are not differentially expressed between parent species																
Lsat_18951	Succinate dehydrogenase subunit 6 mitochondrial	OXPHOS Complex II				X			X		X							
Lsat_10503	succinate dehydrogenase 5	OXPHOS Complex II										X						
TragDub00409-RA	Cytochrome C1 family	OXPHOS Complex III														X		
TragDub16190-RA	photosystem I subunit F	Photosystem I														X		
Lsat_20510	photosystem I subunit O	Photosystem I										X		X	X	X		
Lsat_9599	photosystem II subunit T	Photosystem II		X												X		
TragDub19360-RA	chlorophyll A/B-binding protein	Photosystem II													X	X		
TragDub28661-RA	chlorophyll A/B-binding protein	Photosystem II														X		
Lsat_11580 & Lsat_8208	photosystem II light harvesting complex gene 2.2 & 2.3	Photosystem II														X		
Lsat_32495	cytochrome b6f complex subunit (petM) putative	Cytochrome b6f														X		
TragDub26227-RA	Ribosomal protein L12/ ATP-dependent Clp protease adaptor protein ClpS family protein	Caseinolytic Protease														X		
TragDub25278-RA	sigma factor A	RNA Polymerase PEP														X		
TragDub21893-RA	WHIRLY 1	Dual-Targeted Genes									X							

			Late-flowering group												Early-flowering group			
			Early stage		Early stage					Middle stage					Early vs Mature stage			
			Short-liguled *T. miscellus*					Long-liguled *T. miscellus*		Short-liguled *T. miscellus*				Long-liguled *T. miscellus*	Short-liguled *T. miscellus*			
			3053	3073	104	3049	3053	3073	3066+101	104	3049	3053	3073	3066+101	3053	3073	3053	3073
		No difference in expression b/w polyploid & *T. dubius* but low expression in polyploid than *T. pratensis for* loci that are not differentially expressed between parent species																
Lsat_21648	NADH dehydrogenase	OXPHOS Complex I						X					X					
TragDub20137-RA	Hypoxia-responsive family protein	OXPHOS Complex IV							X									
Lsat_10787	gamma subunit of Mt ATP synthase	OXPHOS Complex V						X				X	X					
TragDub16190-RA	photosystem I subunit F	Photosystem I		X					X				X	X				
Lsat_9332	photosystem II light harvesting complex gene 2.4	Photosystem II											X	X				
Lsat_11580 & Lsat_8208	photosystem II light harvesting complex gene 2.2 & 2.3	Photosystem II												X				
Lsat_32495	cytochrome b6f complex subunit (petM) putative	Cytochrome b6f											X	X				
Lsat_3726	Ribosomal protein PSRP-3/Ycf65	Ribosomal Genes											X					
		No difference in expression b/w polyploid & *T. pratensis* but higher expression in polyploid than *T. dubius* for loci that are not differentially expressed between parent species																
TragDub00693-RA	Ubiquinol-cytochrome C reductase iron-sulfur subunit	OXPHOS Complex III															X	X
TragDub00413-RA	Rubredoxin-like superfamily protein	OXPHOS Complex IV													X	X	X	X
TragDub17408-RA	light-harvesting chlorophyll B-binding protein 3	Photosystem II									X							
Lsat_10456	photosystem II light harvesting complex gene 2.1	Photosystem II															X	X
TragDub22127-RA	plastid-specific ribosomal protein 4	Ribosomal Genes						X										
TragDub00311-RA	sigma factor E	RNA Polymerase PEP															X	
TragDub10196-RA	CBL-interacting protein kinase 14	Dual-Targeted Genes															X	
		No difference in expression b/w polyploid & *T. dubius* but higher expression in polyploid than *T. pratensis* for loci that are not differentially expressed between parent species																
TragDub00409-RA	Cytochrome C1 family	OXPHOS Complex III				X												

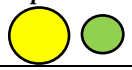
Supplementary Table 3.7: List of genes showing expression level dominance toward either parent species; *Tragopogon pratensis* and *T. dubius* as well as transgressive down-regulation in populations (marked X) of short- and long-liguled *T. miscellus* at early-, middle- and mature-growth stages.

			Late-flowering group											**Early-flowering group**			
			Early stage		**Early stage**					**Middle stage**				**Early vs Mature stage**			
			Short-liguled *T. miscellus*					**Long-liguled *T. miscellus***		**Short-liguled *T. miscellus***			**Long-liguled *T. miscellus***	**Short-liguled *T. miscellus***			
			3053	**3073**	**104**	**3049**	**3053**	**3073**	**3066+101**	**104**	**3049**	**3053**	**3073**	**3066+101**	**3053**	**3073**	**3053**	**3073**
Gene ID	**Gene name**	**Enzyme complex**	**Expression level dominance toward *T. pratensis***															
TragDub00413-RA	Rubredoxin-like superfamily protein	OXPHOS Complex IV	X	X		X			X	X	X	X		X				
Lsat_20510	photosystem I subunit O	Photosystem I	X	X														
TragDub28661-RA	chlorophyll A/B-binding protein	Photosystem II		X					X					X				
Lsat_10456	photosystem II light harvesting complex gene 2.1	Photosystem II								X	X							
TragDub22127-RA	plastid-specific ribosomal protein 4	Ribosomal Genes		X												X		
Lsat_5980	30S ribosomal protein putative	Ribosomal Genes														X		
TragDub00311-RA	sigma factor E	RNA Polymerase PEP									X	X	X					
		Expression level dominance toward *T. dubius*																
TragDub00413-RA	Rubredoxin-like superfamily protein	OXPHOS Complex IV			X			X										
TragDub19360-RA	chlorophyll A/B-binding protein	Photosystem II	X		X	X	X	X		X	X	X	X					
TragDub28661-RA	chlorophyll A/B-binding protein	Photosystem II			X	X	X	X		X	X	X	X					
Lsat_10456	photosystem II light harvesting complex gene 2.1	Photosystem II											X	X				
		Transgressive down-regulation																
Lsat_11580	photosystem II light harvesting complex gene 2.2	Photosystem II											X					
Lsat_8208	photosystem II light harvesting complex gene 2.3	Photosystem II											X					
TragDub00311-RA	sigma factor E	RNA Polymerase PEP														X		

Supplementary Table 3.8: The number and percentage of control genes showing unbiased or biased expression toward either parent species; *Tragopogon dubius or T. pratensis* in each population of short- (sm) and long-liguled (lm) *T. miscellus* at early-, middle- and mature-growth stages.

	Short-liguled *T. miscellus*										**Long-liguled *T. miscellus***		**Average % sm/lm**
	Early stage				**Middle stage**				**Mature stage**		**Early stage**	**Middle stage**	
	104	**3049**	**3053**	**3073**	**104**	**3049**	**3053**	**3073**	**3053**	**3073**	**3066+101**	**3066+101**	
Unbiased 	8 (66.7%)	8 (66.7%)	7 (58.3%)	7 (58.3%)	9 (69.2%)	8 (61.5%)	6 (50%)	7 (53.8%)	9 (69.2%)	5 (38.5%)	9 (75%)	10 (76.9%)	59.2/76.0
Biased toward *T. pratensis* 	1 (8.3%)	3 (25%)	2 (16.7%)	3 (25%)	2 (15.4%)	4 (30.8%)	2 (16.7%)	4 (30.8%)	3 (23.1%)	4 (30.8%)	2 (16.7%)	2 (15.4%)	22.2/16.0
Biased towards *T. dubius* 	3 (25%)	1 (8.3%)	3 (25%)	2 (16.7%)	2 (15.4%)	1 (7.7%)	4 (33.3%)	2 (15.4%)	1 (7.7%)	4 (30.8%)	1 (8.3%)	1 (7.7%)	18.5/8.0
Total	12	12	12	12	13	13	12	13	13	13	12	13	
Biased expression genes	33.3%	33.3%	41.7%	41.7%	30.8%	38.5%	50.0%	46.2%	30.8%	61.6%	25.0%	23.1%	40.8/24.0

Supplementary Table 3.9: The number and percentage of retrograde-signaling partners of dual-targeted genes showing unbiased or biased expression toward either parent species; *Tragopogon dubius* or *T. pratensis* in each population of short- (sm) and long-liguled (lm) *T. miscellus* at early-, middle- and mature-growth stages.

	Short-liguled *T. miscellus*										**Long-liguled *T. miscellus***		**Average % sm/lm**
	Early stage				**Middle stage**				**Mature stage**		**Early stage**	**Middle stage**	
	104	**3049**	**3053**	**3073**	**104**	**3049**	**3053**	**3073**	**3053**	**3073**	**3066+101**	**3066+101**	
Unbiased 	4 (57.1%)	5 (71.4%)	4 (57.1%)	4 (57.1%)	4 (57.1%)	3 (42.9%)	5 (71.4%)	4 (57.1%)	4 (57.1%)	3 (42.9%)	4 (57.1%)	4 (57.1%)	57.1/57.1
Biased toward *T. pratensis* 	1 (14.3%)	0	1 (14.3%)	1 (14.3%)	1 (14.3%)	2 (28.6%)	1 (14.3%)	1 (14.3%)	1 (14.3%)	1 (14.3%)	2 (28.6%)	2 (28.6%)	14.3/28.6
Biased towards *T. dubius* 	2 (28.6%)	2 (28.6%)	2 (28.6%)	2 (28.6%)	2 (28.6%)	2 (28.6%)	1 (14.3%)	2 (28.6%)	2 (28.6%)	3 (42.9%)	1 (14.3%)	1 (14.3%)	28.6/14.3
Total	7	7	7	7	7	7	7	7	7	7	7	7	
Biased expression genes	42.9%	42.9%	42.9%	42.9%	42.9%	57.1%	28.6%	42.9%	42.9%	57.1%	42.9%	42.9%	42.9/42.9

Supplementary Table 3.10: List of homeologs showing differential expression between long-liguled *Tragopogon miscellus* and each population of short-liguled *T. miscellus* at early-growth stage. The color codes represent the genes that show a particular expression pattern across different comparisons. For instance, red color: gene with that specific pattern noticed in all comparisons, green color: in three comparisons, blue color: in any two comparisons and black color: unique to a particular comparison.

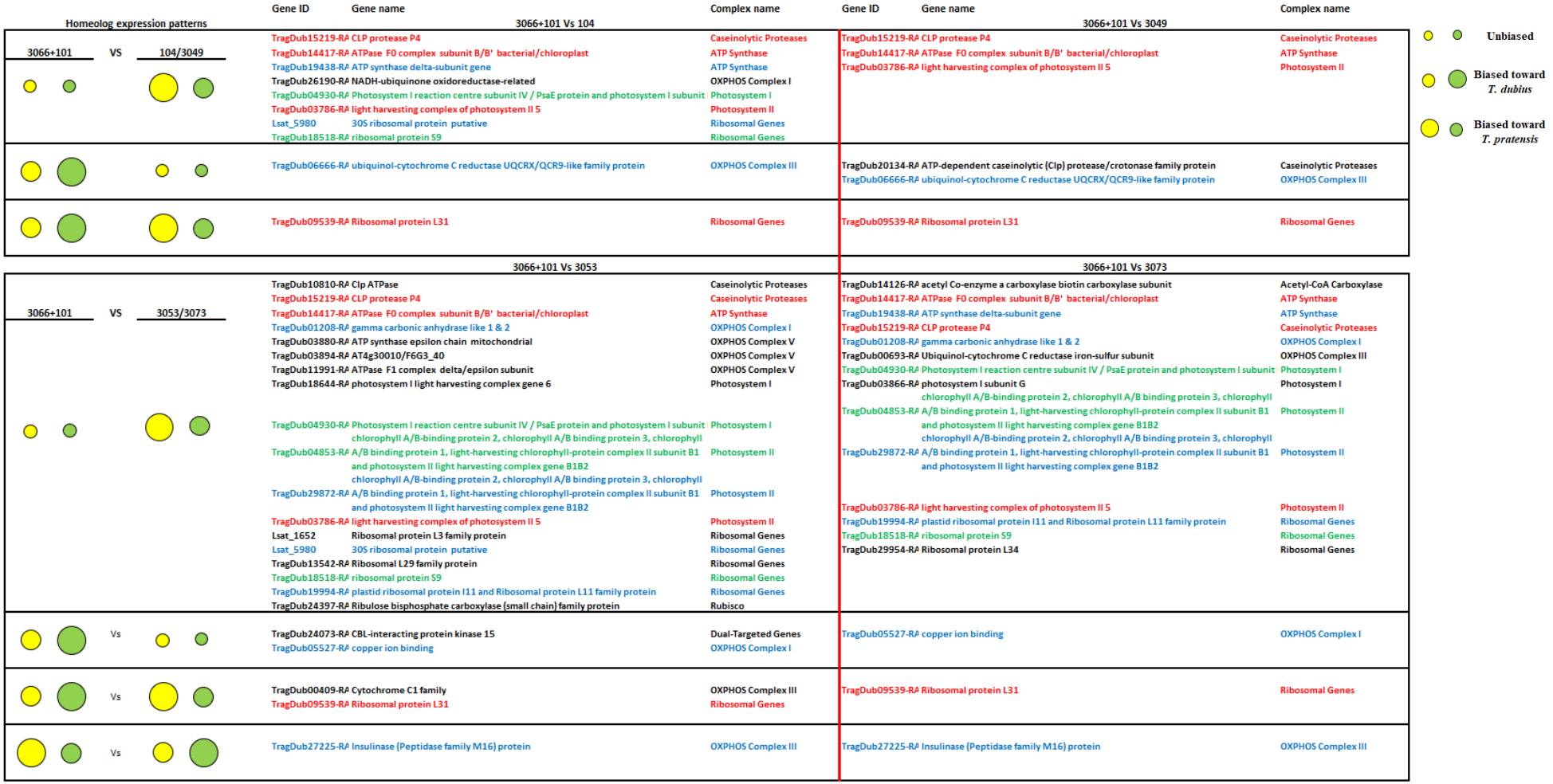
Homeolog expression patterns	Gene ID	Gene name	Complex name	Gene ID	Gene name	Complex name
	3066+101 Vs 104			3066+101 Vs 3049		
3066+101 VS 104/3049	TragDub14417-RA	ATPase F0 complex subunit B/B' bacterial/chloroplast	ATP Synthase	TragDub14417-RA	ATPase F0 complex subunit B/B' bacterial/chloroplast	ATP Synthase
	TragDub19438-RA	ATP synthase delta-subunit gene	ATP Synthase	TragDub15219-RA	CLP protease P4	Caseinolytic Proteases
	TragDub15219-RA	CLP protease P4	Caseinolytic Proteases	TragDub03786-RA	light harvesting complex of photosystem II 5	Photosystem II
	TragDub26190-RA	NADH-ubiquinone oxidoreductase-related	OXPHOS Complex I	TragDub09539-RA	Ribosomal protein L31	Ribosomal Genes
		Photosystem I reaction centre subunit IV / PsaE protein and				
	TragDub04930-RA	photosystem I subunit E-2	Photosystem I			
	TragDub03786-RA	light harvesting complex of photosystem II 5	Photosystem II			
	Lsat_5980	30S ribosomal protein putative	Ribosomal Genes			
	TragDub09539-RA	Ribosomal protein L31	Ribosomal Genes			
	TragDub18518-RA	ribosomal protein S9	Ribosomal Genes			
				TragDub20134-RA	ATP-dependent caseinolytic (Clp) protease/crotonase family protein	Caseinolytic Proteases
				TragDub06666-RA	ubiquinol-cytochrome C reductase UQCRX/QCR9-like family protein	OXPHOS Complex III
	TragDub05540-RA	Chlorophyll A-B binding family protein	Photosystem II	TragDub05540-RA	Chlorophyll A-B binding family protein	Photosystem II
				TragDub02462-RA	ribosomal protein L9	Ribosomal Genes
				TragDub12384-RA	RNApolymerase sigma-subunit F	RNA polymerase PEP
					chlorophyll A/B-binding protein 2, chlorophyll A/B binding protein 3, chlorophyll A/B binding protein 1, light-harvesting chlorophyll-protein complex II subunit B1	
				TragDub04853-RA	and photosystem II light harvesting complex gene B1B2	Photosystem II
	3066+101 Vs 3053			3066+101 Vs 3073		
3066+101 VS 3053/3073	TragDub14417-RA	ATPase F0 complex subunit B/B' bacterial/chloroplast	ATP Synthase	TragDub14126-RA	acetyl Co-enzyme a carboxylase biotin carboxylase subunit	Acetyl-CoA Carboxylase
	TragDub19438-RA	ATP synthase delta-subunit gene	ATP Synthase	TragDub14417-RA	ATPase F0 complex subunit B/B' bacterial/chloroplast	ATP Synthase
	TragDub15219-RA	CLP protease P4	Caseinolytic Proteases	TragDub19438-RA	ATP synthase delta-subunit gene	ATP Synthase
	Lsat_32495	cytochrome b6f complex subunit (petM) putative	Cytochrome b6f	TragDub15219-RA	CLP protease P4	Caseinolytic Proteases
	TragDub01208-RA	gamma carbonic anhydrase like 1 & 2	OXPHOS Complex I	TragDub01208-RA	gamma carbonic anhydrase like 1 & 2	OXPHOS Complex I
	TragDub06877-RA	AT4g30010/F6G3_41	OXPHOS Complex V	TragDub03866-RA	photosystem I subunit G	Photosystem I
	TragDub11991-RA	ATPase F1 complex delta/epsilon subunit	OXPHOS Complex V	TragDub04930-RA	Photosystem I reaction centre subunit IV / PsaE protein and photosystem I subunit I	Photosystem I
		Photosystem I reaction centre subunit IV / PsaE protein and			chlorophyll A/B-binding protein 2, chlorophyll A/B binding protein 3, chlorophyll A/B binding protein 1, light-harvesting chlorophyll-protein complex II subunit B1	
	TragDub04930-RA	photosystem I subunit E-2	Photosystem I	TragDub29872-RA	and photosystem II light harvesting complex gene B1B2	Photosystem II
	TragDub18644-RA	photosystem I light harvesting complex gene 6	Photosystem I	TragDub03786-RA	light harvesting complex of photosystem II 5	Photosystem II
	TragDub29220-RA	photosystem I light harvesting complex gene 3	Photosystem I	Lsat_5980	30S ribosomal protein putative	Ribosomal Genes
	TragDub03786-RA	light harvesting complex of photosystem II 5	Photosystem II	TragDub19994-RA	plastid ribosomal protein l11 and Ribosomal protein L11 family protein	Ribosomal Genes
	Lsat_5980	30S ribosomal protein putative	Ribosomal Genes	TragDub29954-RA	Ribosomal protein L34	Ribosomal Genes
	TragDub13542-RA	Ribosomal L29 family protein	Ribosomal Genes	TragDub09539-RA	Ribosomal protein L31	Ribosomal Genes
	TragDub18518-RA	ribosomal protein S9	Ribosomal Genes	TragDub18518-RA	ribosomal protein S9	Ribosomal Genes
	TragDub19994-RA	plastid ribosomal protein l11 and Ribosomal protein L11 family protein	Ribosomal Genes	TragDub24397-RA	Ribulose bisphosphate carboxylase (small chain) family protein	Rubisco
	TragDub24397-RA	Ribulose bisphosphate carboxylase (small chain) family protein	Rubisco			
	TragDub05527-RA	copper ion binding	OXPHOS Complex I	TragDub26227-RA	Ribosomal protein L12/ ATP-dependent Clp protease adaptor protein ClpS family pr	Caseinolytic Proteases
	TragDub00409-RA	Cytochrome C1 family	OXPHOS Complex III	TragDub05527-RA	copper ion binding	OXPHOS Complex I
	TragDub05540-RA	Chlorophyll A-B binding family protein	Photosystem II	TragDub00409-RA	Cytochrome C1 family	OXPHOS Complex III
	Lsat_1652	Ribosomal protein L3 family protein	Ribosomal Genes	TragDub05540-RA	Chlorophyll A-B binding family protein	Photosystem II

Unbiased

Biased toward *T. dubius*

Biased toward *T. pratensis*

Supplementary Table 3.11: List of homeologs showing differential expression between long-liguled *Tragopogon miscellus* and each population of short-liguled *T. miscellus* at middle-growth stage. The color codes represent the genes that show a particular expression pattern across different comparisons. For instance, red color: gene with that specific pattern noticed in all comparisons, green color: in three comparisons, blue color: in any two comparisons and black color: unique to a particular comparison.



Homeolog expression patterns	Gene ID	Gene name	3066+101 Vs 104 Complex name	Gene ID	Gene name	3066+101 Vs 3049 Complex name
3066+101 VS 104/3049	TragDub15219-RA	CLP protease P4	Caseinolytic Proteases	TragDub15219-RA	CLP protease P4	Caseinolytic Proteases
	TragDub14417-RA	ATPase F0 complex subunit B/B' bacterial/chloroplast	ATP Synthase	TragDub14417-RA	ATPase F0 complex subunit B/B' bacterial/chloroplast	ATP Synthase
	TragDub19438-RA	ATP synthase delta-subunit gene	ATP Synthase	TragDub03786-RA	light harvesting complex of photosystem II 5	Photosystem II
	TragDub26190-RA	NADH-ubiquinone oxidoreductase-related	OXPHOS Complex I			
	TragDub04930-RA	Photosystem I reaction centre subunit IV / PsaE protein and photosystem I subunit	Photosystem I			
	TragDub03786-RA	light harvesting complex of photosystem II 5	Photosystem II			
	Lsat_5980	30S ribosomal protein putative	Ribosomal Genes			
	TragDub18518-RA	ribosomal protein S9	Ribosomal Genes			
	TragDub06666-RA	ubiquinol-cytochrome C reductase UQCRX/QCR9-like family protein	OXPHOS Complex III	TragDub20134-RA	ATP-dependent caseinolytic (Clp) protease/crotonase family protein	Caseinolytic Proteases
				TragDub06666-RA	ubiquinol-cytochrome C reductase UQCRX/QCR9-like family protein	OXPHOS Complex III
	TragDub09539-RA	Ribosomal protein L31	Ribosomal Genes	TragDub09539-RA	Ribosomal protein L31	Ribosomal Genes

Homeolog expression patterns	Gene ID	Gene name	3066+101 Vs 3053 Complex name	Gene ID	Gene name	3066+101 Vs 3073 Complex name
3066+101 VS 3053/3073	TragDub10810-RA	Clp ATPase	Caseinolytic Proteases	TragDub14126-RA	acetyl Co-enzyme a carboxylase biotin carboxylase subunit	Acetyl-CoA Carboxylase
	TragDub15219-RA	CLP protease P4	Caseinolytic Proteases	TragDub14417-RA	ATPase F0 complex subunit B/B' bacterial/chloroplast	ATP Synthase
	TragDub14417-RA	ATPase F0 complex subunit B/B' bacterial/chloroplast	ATP Synthase	TragDub19438-RA	ATP synthase delta-subunit gene	ATP Synthase
	TragDub01208-RA	gamma carbonic anhydrase like 1 & 2	OXPHOS Complex I	TragDub15219-RA	CLP protease P4	Caseinolytic Proteases
	TragDub03880-RA	ATP synthase epsilon chain mitochondrial	OXPHOS Complex V	TragDub01208-RA	gamma carbonic anhydrase like 1 & 2	OXPHOS Complex I
	TragDub03894-RA	AT4g30010/F6G3_40	OXPHOS Complex V	TragDub00693-RA	Ubiquinol-cytochrome C reductase iron-sulfur subunit	OXPHOS Complex III
	TragDub11991-RA	ATPase F1 complex delta/epsilon subunit	OXPHOS Complex V	TragDub04930-RA	Photosystem I reaction centre subunit IV / PsaE protein and photosystem I subunit	Photosystem I
	TragDub18644-RA	photosystem I light harvesting complex gene 6	Photosystem I	TragDub03866-RA	photosystem I subunit G	Photosystem I
	TragDub04930-RA	Photosystem I reaction centre subunit IV / PsaE protein and photosystem I subunit chlorophyll A/B-binding protein 2, chlorophyll A/B binding protein 3, chlorophyll	Photosystem I		chlorophyll A/B-binding protein 2, chlorophyll A/B binding protein 3, chlorophyll	
	TragDub04853-RA	A/B binding protein 1, light-harvesting chlorophyll-protein complex II subunit B1 and photosystem II light harvesting complex gene B1B2 chlorophyll A/B-binding protein 2, chlorophyll A/B binding protein 3, chlorophyll	Photosystem II	TragDub04853-RA	A/B binding protein 1, light-harvesting chlorophyll-protein complex II subunit B1 and photosystem II light harvesting complex gene B1B2 chlorophyll A/B-binding protein 2, chlorophyll A/B binding protein 3, chlorophyll	Photosystem II
	TragDub29872-RA	A/B binding protein 1, light-harvesting chlorophyll-protein complex II subunit B1 and photosystem II light harvesting complex gene B1B2	Photosystem II	TragDub29872-RA	A/B binding protein 1, light-harvesting chlorophyll-protein complex II subunit B1 and photosystem II light harvesting complex gene B1B2	Photosystem II
	TragDub03786-RA	light harvesting complex of photosystem II 5	Photosystem II	TragDub03786-RA	light harvesting complex of photosystem II 5	Photosystem II
	Lsat_1652	Ribosomal protein L3 family protein	Ribosomal Genes	TragDub19994-RA	plastid ribosomal protein l11 and Ribosomal protein L11 family protein	Ribosomal Genes
	Lsat_5980	30S ribosomal protein putative	Ribosomal Genes	TragDub18518-RA	ribosomal protein S9	Ribosomal Genes
	TragDub13542-RA	Ribosomal L29 family protein	Ribosomal Genes	TragDub29954-RA	Ribosomal protein L34	Ribosomal Genes
	TragDub18518-RA	ribosomal protein S9	Ribosomal Genes			
	TragDub19994-RA	plastid ribosomal protein l11 and Ribosomal protein L11 family protein	Ribosomal Genes			
	TragDub24397-RA	Ribulose bisphosphate carboxylase (small chain) family protein	Rubisco			
Vs	TragDub24073-RA	CBL-interacting protein kinase 15	Dual-Targeted Genes	TragDub05527-RA	copper ion binding	OXPHOS Complex I
	TragDub05527-RA	copper ion binding	OXPHOS Complex I			
Vs	TragDub00409-RA	Cytochrome C1 family	OXPHOS Complex III	TragDub09539-RA	Ribosomal protein L31	Ribosomal Genes
	TragDub09539-RA	Ribosomal protein L31	Ribosomal Genes			
Vs	TragDub27225-RA	Insulinase (Peptidase family M16) protein	OXPHOS Complex III	TragDub27225-RA	Insulinase (Peptidase family M16) protein	OXPHOS Complex III

Supplementary Table 3.12: List of loci differentially expressed across growth stages in populations of short- and long-liguled *Tragopogon miscellus*. H2>H1: *T. pratensis* bias, H2<H1: *T. dubius* bias and H2~H1: unbiased expression. Filtered out means that gene was filtered. out as a result of filtering criteria.

				Late-flowering group										Early-flowering group			
				104		3049		3053		3073		3066+101		3053		3073	
Sr. #	Gene ID	Gene name	Enzyme complex	Early-age	Middle-age	Early-age	Middle-age	Early-age	Middle-age	Early-age	Middle-age	Early-age	Middle-age	Early-age	Middle-age	Early-age	Middle-age
1	TragDub02230-RA	chloroplastic acetylcoenzyme A carboxylase 1 and biotin carboxyl carrier protein 2	Acetyl-CoA Carboxylase									H2>H1	H2~H1				
2	TragDub14126-RA	acetyl Co-enzyme a carboxylase biotin carboxylase subunit	Acetyl-CoA Carboxylase													H2~H1	H2>H1
3	TragDub14823-RA	CLPT1	Caseinolytic Proteases									H2<H1	H2~H1				
4	Lsat_32495	cytochrome b6f complex subunit (petM) putative	Cytochrome b6f									H2~H1	H2>H1				
5	TragDub24280-RA	photosynthetic electron transfer C	Cytochrome b6f													H2<H1	H2~H1
6	TragDub26682-RA	PPR protein localized to the nucleus and mitochondria 1	Dual-Targeted Genes			H2~H1	H2>H1										
7	TragDub01137-RA	Mitochondrial transcription termination factor 6	Dual-Targeted Genes							H2~H1	H2<H1						
8	TragDub04484-RA	NADH-ubiquinone oxidoreductase 20 kDa subunit mitochondrial	OXPHOS Complex I	H2~H1	H2<H1												
9	TragDub11786-RA	At1g67350; NADH-ubiquinone oxidoreductase	OXPHOS Complex I	H2~H1	H2<H1												
10	TragDub24868-RA	L-galactono-1 4-lactone dehydrogenase	OXPHOS Complex I									H2~H1	H2>H1				
11	TragDub16379-RA	NADH-ubiquinone oxidoreductase B8 subunit	OXPHOS Complex I											H2~H1	H2>H1		
12	TragDub14659-RA	LYR family of Fe/S cluster biogenesis protein	OXPHOS Complex I											H2>H1	H2~H1		
13	TragDub23106-RA	Mitochondrial import inner membrane translocase subunit Tim17/Tim22/Tim23 family protein	OXPHOS Complex I													H2<H1	H2~H1
14	TragDub07138-RA	NADH dehydrogenase	OXPHOS Complex I													H2~H1	H2<H1
15	TragDub06666-RA	ubiquinol-cytochrome C reductase UQCRX/QCR9-like family protein	OXPHOS Complex III	H2<H1	H2~H1												
16	TragDub14875-RA	Ubiquinol-cytochrome C reductase iron-sulfur subunit	OXPHOS Complex III	H2<H1	H2~H1	H2<H1	H2~H1							H2<H1	H2~H1		
17	TragDub00693-RA	Ubiquinol-cytochrome C reductase iron-sulfur subunit	OXPHOS Complex III													H2~H1	H2>H1
18	TragDub29090-RA	Protein of unknown function (DUF1640)	OXPHOS Complex IV											H2<H1	H2~H1		
19	TragDub00413	Rubredoxin-like superfamily protein	OXPHOS Complex IV											filtered out	H2<H1		
20	TragDub17478-RA	cytochrome C oxidase 6B	OXPHOS Complex IV													H2<H1	H2~H1
21	TragDub06877-RA	AT4g30010/F6G3_41	OXPHOS Complex V					H2>H1	H2~H1								
22	Lsat_10787	gamma subunit of Mt ATP synthase	OXPHOS Complex V					H2>H1	filtered out							H2~H1	H2>H1
23	TragDub13305-RA	ATP synthase D chain mitochondrial	OXPHOS Complex V									H2~H1	H2<H1				
24	TragDub16190-RA	photosystem I subunit F	Photosystem I	H2>H1	filtered out												
25	TragDub29220-RA	photosystem I light harvesting complex gene 3	Photosystem I					H2>H1	H2~H1								
26	TragDub10108-RA	photosystem I subunit H-1 and photosystem I subunit H2	Photosystem I							H2~H1	H2>H1					H2~H1	H2<H1
27	TragDub18644-RA	photosystem I light harvesting complex gene 6	Photosystem I											H2~H1	H2>H1		
28	TragDub29872-RA	chlorophyll A/B-binding protein/photosystem II light harvesting complex gene	Photosystem II					H2~H1	H2>H1								
29	TragDub28226-RA	chlorophyll A/B-binding protein/photosystem II light harvesting complex gene	Photosystem II							H2~H1	H2<H1						
30	Lsat_11580	photosystem II light harvesting complex gene 2.2	Photosystem II							H2~H1	H2>H1			H2>H1	H2~H1		
31	TragDub03856-RA	photosystem II subunit R	Photosystem II									H2<H1	H2~H1				
32	TragDub04853-RA	chlorophyll A/B-binding protein/photosystem II light harvesting complex gene	Photosystem II									H2>H1	H2~H1				
33	TragDub05540-RA	Chlorophyll A-B binding family protein	Photosystem II									H2<H1	H2~H1				
34	Lsat_8208	photosystem II light harvesting complex gene 2.3	Photosystem II											H2>H1	H2~H1	H2~H1	H2>H1
35	TragDub21729-RA	ribosomal protein L4	Ribosomal Genes	H2~H1	H2<H1			H2<H1	H2~H1								
36	Lsat_17875	Ribosomal L5P family protein	Ribosomal Genes	H2>H1	filtered out									H2~H1	filtered out	H2>H1	filtered out
37	TragDub21350-RA	Ribosomal protein L19 family protein	Ribosomal Genes					H2<H1	H2~H1	H2~H1	H2<H1						
38	TragDub18778-RA	ribosomal protein large subunit 27	Ribosomal Genes			H2~H1	H2>H1	H2>H1	H2~H1	H2~H1	H2>H1						
39	TragDub10107-RA	ribosomal protein S17	Ribosomal Genes	H2<H1	H2~H1			H2<H1	H2~H1								
40	TragDub02462-RA	ribosomal protein L9	Ribosomal Genes			H2~H1	H2<H1										
41	TragDub26728-RA	Ribosomal L28 family	Ribosomal Genes			H2~H1	H2>H1										
42	TragDub16283-RA	Ribosomal L18p/L5e family protein	Ribosomal Genes			H2~H1	H2<H1							H2<H1	H2~H1		
43	TragDub03561-RA	ribosomal protein L15	Ribosomal Genes							H2~H1	H2<H1						
44	TragDub09539-RA	Ribosomal protein L31	Ribosomal Genes									H2~H1	H2<H1				
45	Lsat_1652	Ribosomal protein L3 family protein	Ribosomal Genes									H2<H1	H2~H1	H2<H1	H2~H1		
46	TragDub11958-RA	Translation protein SH3-like family protein	Ribosomal Genes													H2~H1	H2<H1
47	TragDub12384-RA	RNApolymerase sigma-subunit F	RNA polymerase PEP			H2~H1	H2<H1										
48	TragDub24770	RNApolymerase sigma subunit 2	RNA polymerase PEP									filtered out	H2<H1				

Supplementary Table 3.13: Number of organellar genes differentially expressed between diploid parent species, *Tragopogon dubius* and *T. pratensis* at different growth stages.

Expression patterns	Early stage	Middle stage	Mature stage	Earl stage-late flowering plants	Early stage-early flowering plants
No differential expression	77	77	76	77	80
Expression higher in *T. pratensis*	2	2	0	3	0
Expression higher in *T. dubius*	2	2	5	1	1
Total	81	81	81	81	81

Supplementary Table 3.14: Number of organellar genes differentially expressed between long-liguled *Tragopogon miscellus* (lm) and populations of short-liguled *T. miscellus* (sm) at different growth stages.

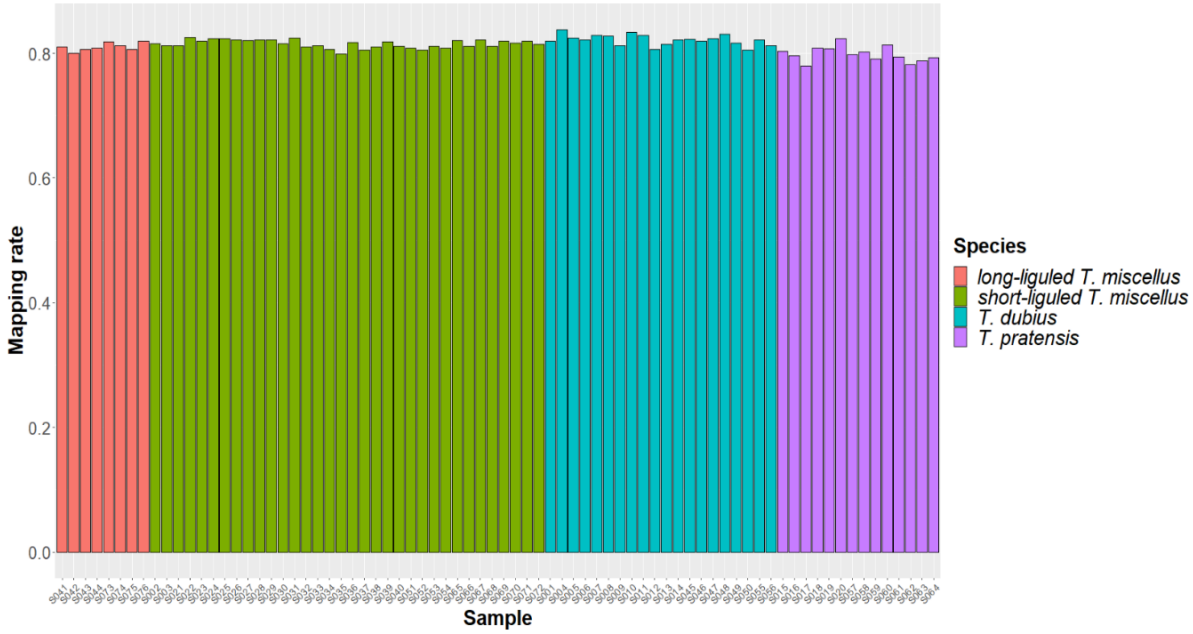
	104		3049		3053		3073	
Expression patterns	**Early stage**	**Middle stage**	**Early stage**	**Middle stage**	**Early stage**	**Middle stage**	**Early stage**	**Middle stage**
No differential expression	76	77	79	77	78	78	79	79
Expression higher in sm	0	2	0	3	0	1	0	1
Expression higher in lm	5	2	2	1	3	2	2	1
Total	81	81	81	81	81	81	81	81

Supplementary Table 3.15: Number of organellar genes differentially expressed between populations of short- and long-liguled *Tragopogon miscellus* and their respective maternal parents, at different growth stages.

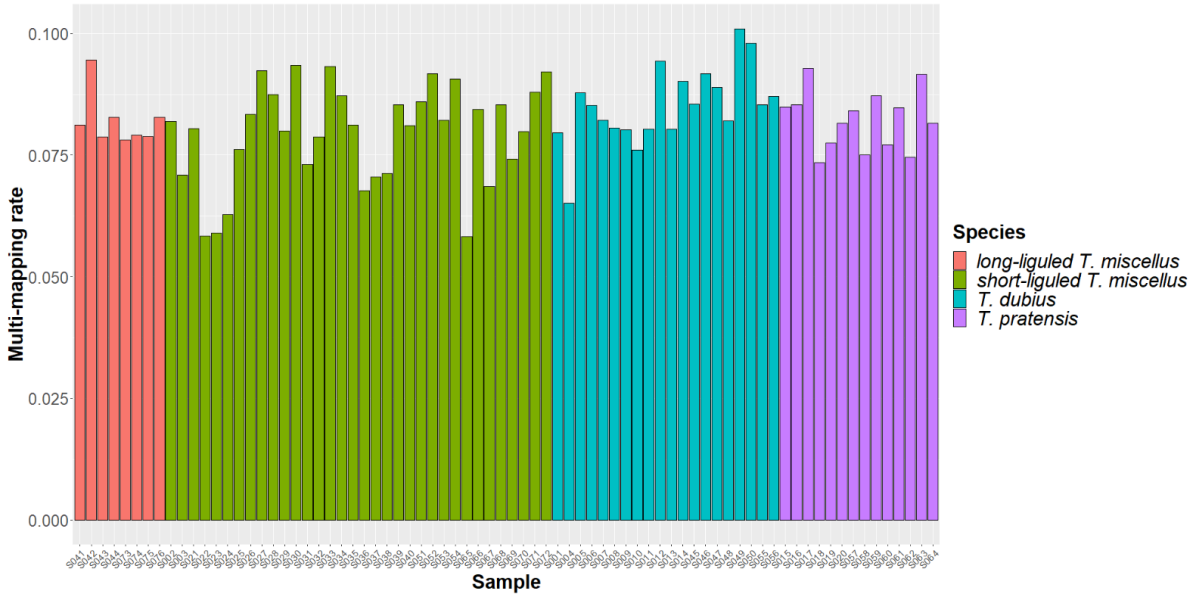
	Short-liguled *T. miscellus* (sm) Vs *T. pratensis*														**Long-liguled *T. miscellus* (lm) Vs *T. dubius***	
	104		**3049**		**3053**					**3073**					**3066+101**	
Expression patterns	**Early stage**	**Middle stage**	**Early stage**	**Middle stage**	**Early stage**	**Middle stage**	**Mature stage**	**Early stage-late flowering plants**	**Early stage-early flowering plants**	**Early stage**	**Middle stage**	**Mature stage**	**Early stage-late flowering plants**	**Early stage-early flowering plants**	**Early stage**	**Middle stage**
No differential expression	73	78	80	80	80	79	76	77	81	80	81	80	80	74	80	80
Expression higher in *T. pratensis*/*T.dubius*	7	3	0	1	0	1	0	3	0	0	0	0	0	6	0	0
Expression higher in sm/lm	0	0	0	0	0	0	4	0	0	0	0	0	0	0	1	1
Total	80	81	80	81	80	80	80	80	81	80	81	80	80	80	81	81

Supplementary Table 3.16: List of organelle genes differentially expressed between parent species, between short- and long-liguled *Tragopogon miscellus* as well as between both reciprocal forms and their respective maternal parents.

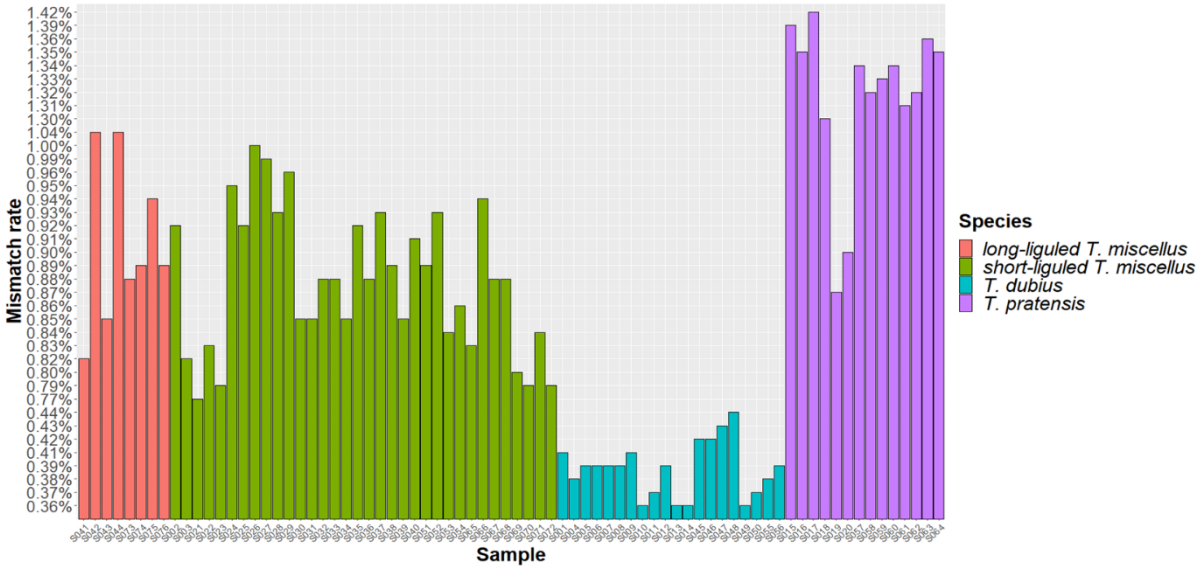
				Late-flowering group																				Early-flowering gorup					
				T. dubius (P1) Vs *T. pratensis* (P2)		Long-liguled *T. miscellus* (lm) Vs Short-liguled *T. miscellus* (sm)								*T. Pratensis* (P2) Vs sm								*T.dubius* (P1) Vs lm		*T. dubius* (P1) Vs *T. pratensis* (P2)		*T. Pratensis* (P2) Vs sm			
						lm vs 104		lm vs 3049		lm vs 3053		lm vs 3073		P2 vs 104		P2 vs 3049		P2 vs 3053		P2 vs 3073						P2 vs 3053		P2 vs 3073	
Gene ID	Gene name	Genome	Enzyme complex	Early-age	Middle-age	Early-age	Middle-age	Early-age	Middle-age	Early-age	Middle-age	Early-age	Middle-age	Early-age	Middle-age	Early-age	Middle-age	Early-age	Middle-age	Early-age	Middle-age	Early-age	Middle-age	Early-age	Mature-age	Early-age	Mature-age	Early-age	Mature-age
L_sativa_-_atp1A_gene	ATPase subunit 1	Mitochondiral	OXPHOS Complex V	P2~P1	P2<P1	lm>sm	lm~sm	lm>sm	lm~sm	lm>sm	lm>sm	lm>sm	lm~sm																
L_sativa_-_atp1B_gene	ATPase subunit 1	Mitochondiral	OXPHOS Complex V	P2<P1	P2<P1	lm>sm	lm>sm	lm>sm	lm>sm	lm>sm	lm>sm	lm>sm	lm>sm	P2>sm	P2~sm									P2<P1	P2<P1				
L_sativa_-_atp4_gene	ATPase subunit 4	Mitochondiral	OXPHOS Complex V			lm~sm	lm>sm																						
L_sativa_-_atp9_gene	ATPase subunit 9	Mitochondiral	OXPHOS Complex V	P2>P1	P2~P1	lm~sm	lm<sm							P2>sm	P2~sm											P2~sm	P2<sm		
L_sativa_-_cox2_gene	Cytochrome c oxidase subunit 2	Mitochondiral	OXPHOS Complex IV	P2>P1	P2>P1									P2~sm	P2>sm							P1<lm	P1~lm						
L_sativa_-_cob_gene	Apocytochrome b	Mitochondiral	OXPHOS Complex III											P2~sm	P2>sm														
L_sativa_-_nad3_gene	NADH dehydrogenase subunit 3	Mitochondiral	OXPHOS Complex I											P2~sm	P2>sm	P2~sm	P2>sm												
L_sativa_-_nad6_gene	NADH dehydrogenase subunit 6	Mitochondiral	OXPHOS Complex I	P2>P1	P2>P1									P2>sm	P2~sm			P2>sm	P2>sm			P1~lm	P1<lm			P2~sm	P2<sm		
trago_dubi_-_petG_gene	Cytochrome b6/f complex subunit	Chloroplast	Cytochrome b6f									lm~sm	lm<sm	P2>sm	P2~sm			P2>sm	P2~sm							P2~sm	P2<sm	P2>sm	P2~sm
trago_dubi_-_petL_gene	Cytochrome b6/f complex subunit	Chloroplast	Cytochrome b6f															P2>sm	P2~sm										
trago_dubi_-_psaJ_gene	Photosystem I subunit IX	Chloroplast	photosystem I			lm~sm	lm<sm	lm~sm	lm<sm																				
trago_dubi_-_psbA_gene	Photosystem II protein D1	Chloroplast	photosystem II					lm~sm	lm<sm																				
trago_dubi_-_psbT_gene	Photosystem II protein T	Chloroplast	photosystem II											P2>sm	P2~sm									P2~P1	P2<P1				
trago_dubi_-_psbK_gene	Photosystem II protein K	Chloroplast	photosystem II																									P2>sm	P2~sm
trago_dubi_-_psbZ_gene	Photosystem II protein Z	Chloroplast	photosystem II																									P2>sm	P2~sm
trago_dubi_-_rbcL_gene	carboxylase/oxygenase large subunit	Chloroplast	Rubisco											P2>sm	P2~sm														
trago_dubi_-_rpl32_gene	Ribosomal protein L32	Chloroplast	Ribosomal Genes			lm>sm	lm~sm			lm>sm	lm~sm			P2>sm	P2~sm									P2~P1	P2<P1			P2>sm	P2~sm
trago_dubi_-_rps3_gene	Ribosomal protein S3	Chloroplast	Ribosomal Genes																							P2~sm	P2<sm		
trago_dubi_-_atpB_gene	ATP synthase CF1 beta subunit	Chloroplast	ATP Synthase			lm >sm	lm~sm																	P2~P1	P2<P1			P2>sm	P2~sm
trago_dubi_-_atpE_gene	ATP synthase CF1 epsilon subunit	Chloroplast	ATP Synthase			lm >sm	lm~sm																	P2~P1	P2<P1			P2>sm	P2~sm
trago_dubi_-_atpH_gene	ATP synthase CF0 C subunit	Chloroplast	ATP Synthase					lm~sm	lm<sm	lm~sm	lm<sm																		



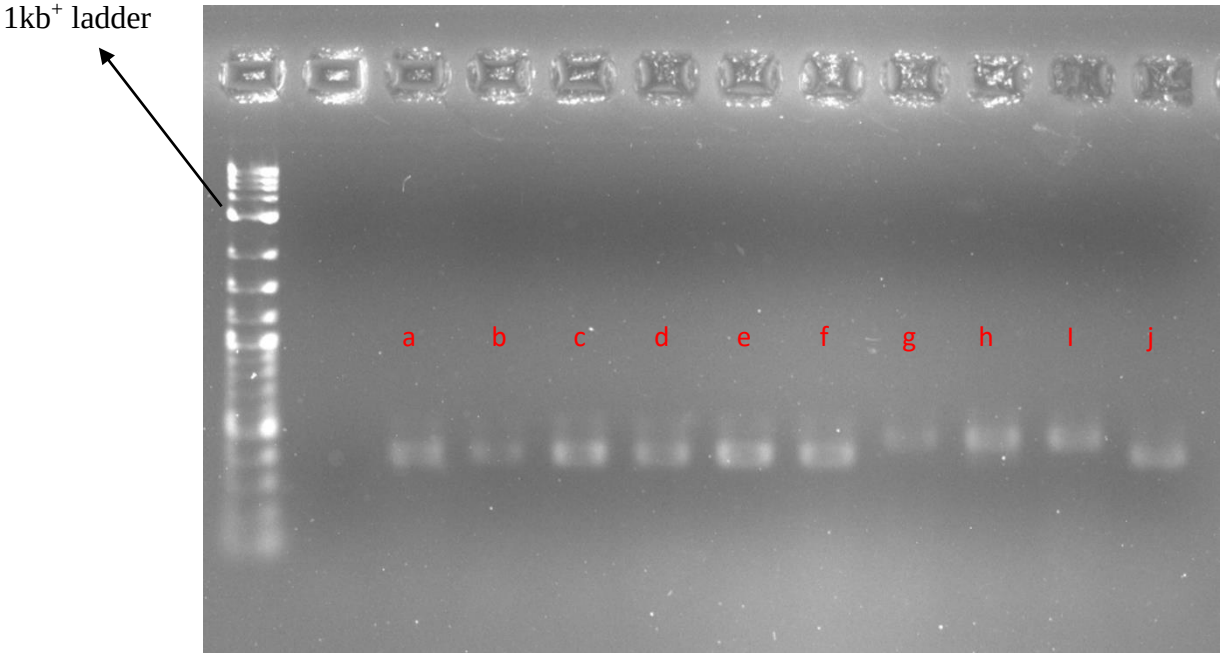
Supplementary Figure 3.1: Mapping rate of raw reads mapped to the reference gene models in all examined samples of *Tragopogon dubius* (blue), *T. pratensis* (purple) and reciprocal forms (orange: long-liguled, green: short-liguled) of *T. miscellus* allopolyploids.



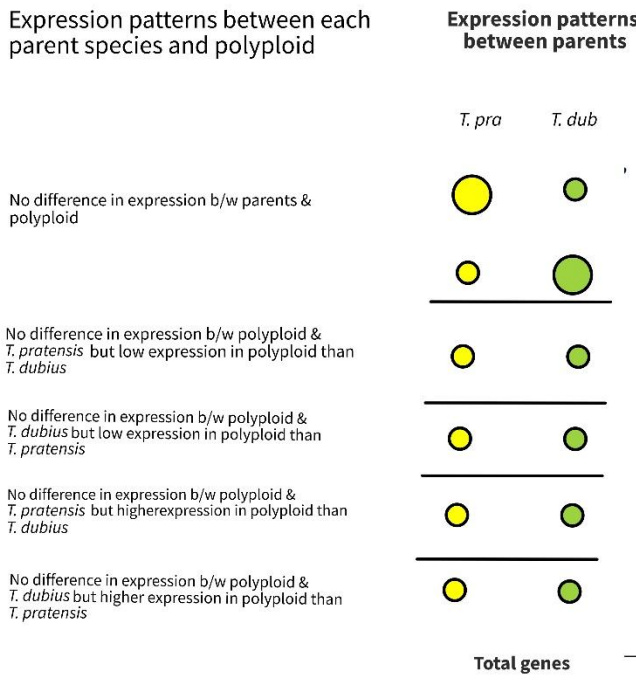
Supplementary Figure 3.2: Multi-mapping rate demonstrating the percentage of reads mapped to multiple reference gene models in all individual samples of *Tragopogon dubius* (blue), *T. pratensis* (purple) and reciprocal forms (orange: long-liguled, green: short-liguled) of *T. miscellus* allopolyploids.



Supplementary Figure 3.3: Mean proportion of mismatches between raw reads and reference gene models in all individual samples of *Tragopogon dubius* (blue), *T. pratensis* (purple) and reciprocal forms (orange: long-liguled, green: short-liguled) of *T. miscellus* allopolyploids.

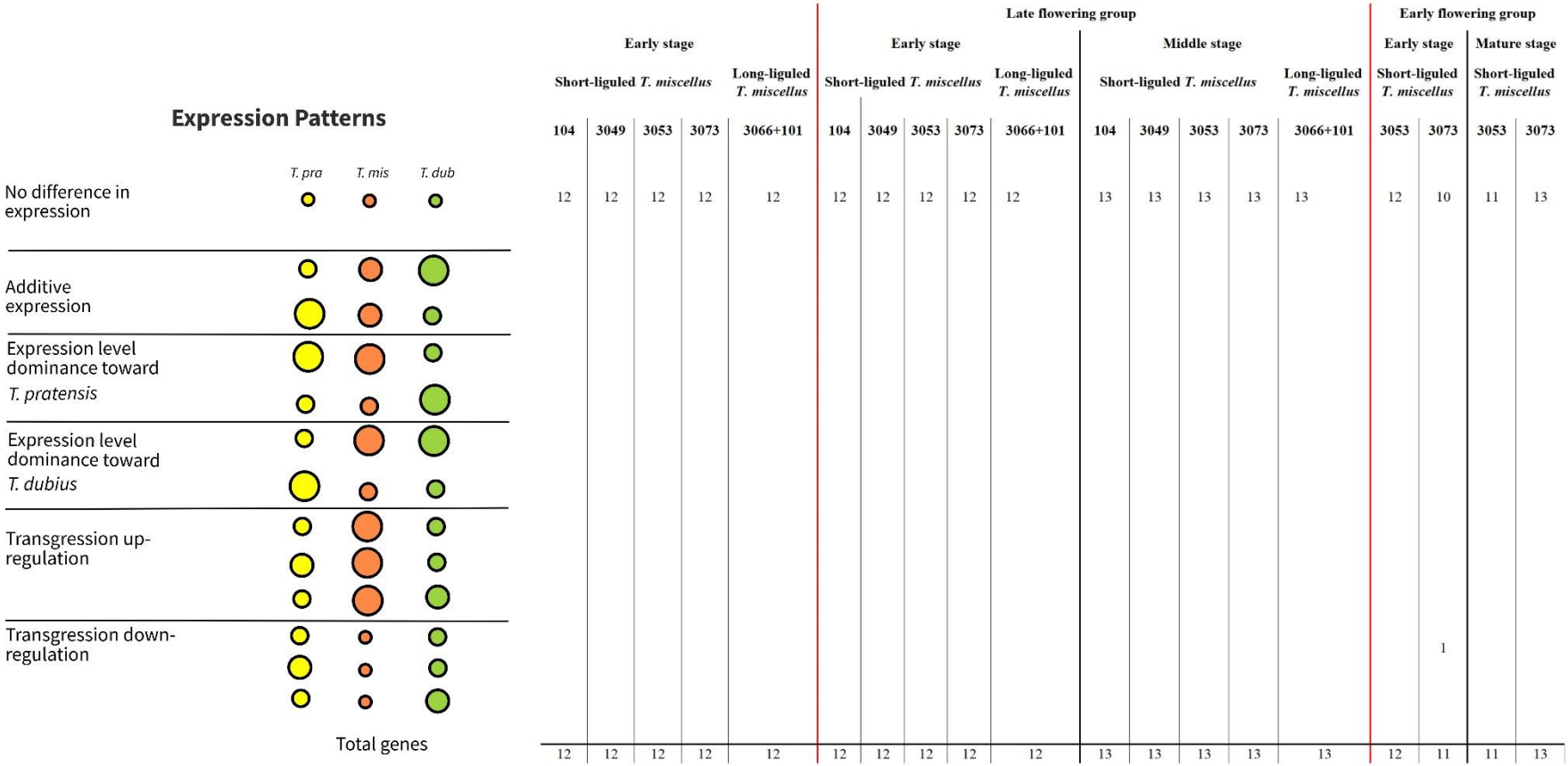


Supplementary Figure 3.4: PCR confirmation of the ploidy level of diploid parent species via TDF-85 gene amplification. Bands a-f and j correspond to various *Tragopogon dubius* individuals and g-i bands show different *T. pratensis* individuals. h is the *T. pratensis* individual with suspicious behavior, clustered with allopolyploid species in principal component analysis.

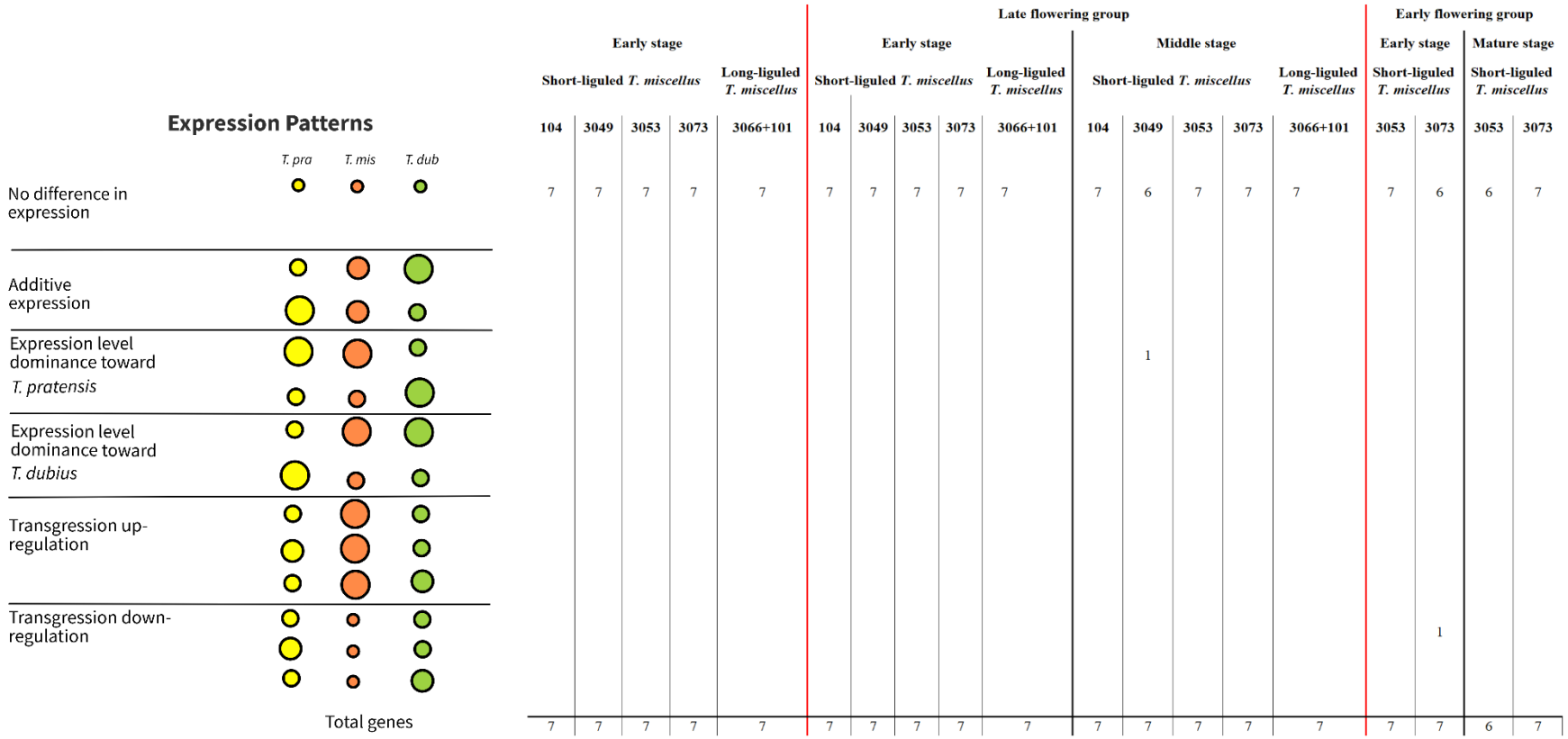


	Late flowering group															Early flowering group			
	Early stage					Early stage					Middle stage					Early stage		Mature stage	
	Short-liguled *T. miscellus*				Long-liguled *T. miscellus*	Short-liguled *T. miscellus*				Long-liguled *T. miscellus*	Short-liguled *T. miscellus*				Long-liguled *T. miscellus*	Short-liguled *T. miscellus*		Short-liguled *T. miscellus*	
	104	3049	3053	3073	3066+101	104	3049	3053	3073	3066+101	104	3049	3053	3073	3066+101	3053	3073	3053	3073
No difference in expression b/w parents & polyploid	1	1	1	0	1	1	1	1	1	1	3	1	3	2	3	1	0	0	0
	2	2	1	1	1	2	2	2	2	2	1	1	1	1	1	1	0	1	1
No difference in expression b/w polyploid & *T. pratensis* but low expression in polyploid than *T. dubius*	1	0	0	1	1	0	1	0	0	1	0	2	0	2	1	2	11	0	0
No difference in expression b/w polyploid & *T. dubius* but low expression in polyploid than *T. pratensis*	0	0	0	1	2	0	0	0	2	2	0	0	1	6	5	0	0	0	0
No difference in expression b/w polyploid & *T. pratensis* but higherexpression in polyploid than *T. dubius*	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	1	5	3
No difference in expression b/w polyploid & *T. dubius* but higher expression in polyploid than *T. pratensis*	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Total genes	4	3	2	3	5	3	5	3	6	6	4	5	5	11	10	5	12	6	4

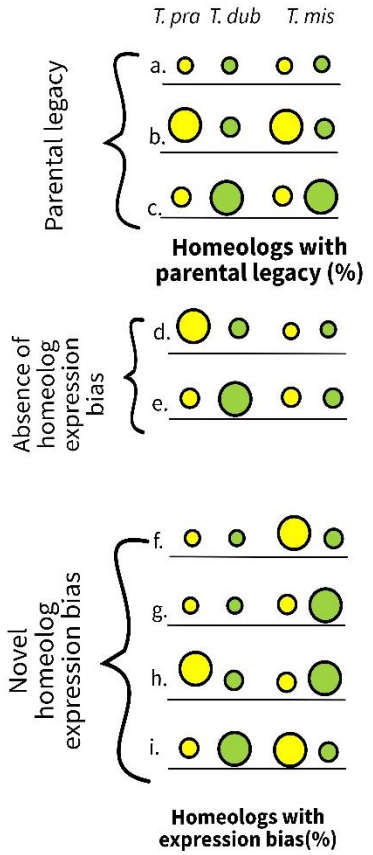
Supplementary Figure 3.5: Number of genes showing anomalous expression patterns in reciprocal forms of *Tragopogon miscellus* allopolyploids compared to their parent species. These genes did not fall into any expression categories mentioned in Figure 3.7.



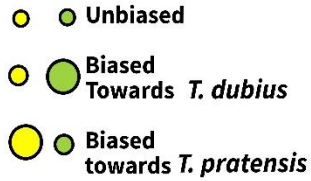
Supplementary Figure 3.6: Comparison of expression patterns of 12-13 duplicated control genes in reciprocal forms of *Tragopogon miscellus* allopolyploids relative to their diploid parent species. *T. pra*, *T. dub* and *T. mis* = *T. pratensis*, *T. dubius* and *T. miscellus*, respectively. First comparison at early stage was done by combining the early age samples of both late and early flowering plants.



Supplementary Figure 3.7: Comparison of expression patterns of seven duplicated retrograde-signaling partners of dual-targeted genes in reciprocal forms of *Tragopogon miscellus* allopolyploids relative to their diploid parent species. *T. pra*, *T. dub* and *T. mis* = *T. pratensis*, *T. dubius* and *T. miscellus*, respectively. First comparison at early stage was done by combining the early age samples of both late and early flowering plants.

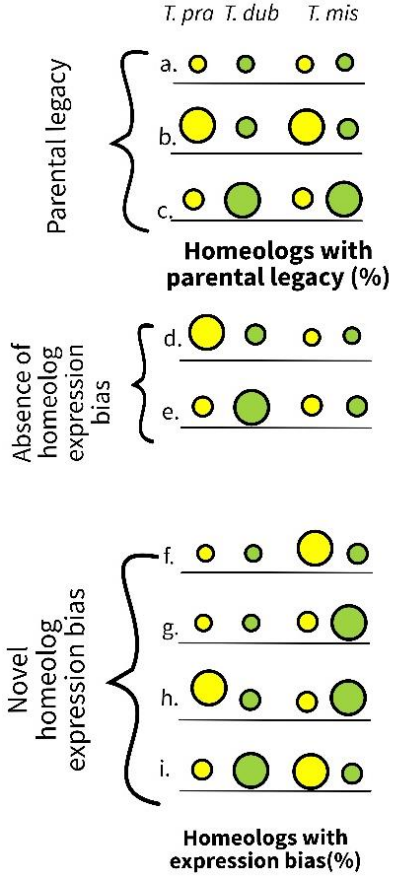


	Late flowering gorup											Early flowering group			
	Early stage		Early stage					Middle stage					Early vs Mature stage			
	Short-liguled *T. miscellus*						Long-liguled *T. miscellus*	Short-liguled *T. miscellus*				Long-liguled *T. miscellus*	Short-liguled *T. miscellus*			
	3053*	3073*	104	3049	3053	3073	3066+101	104	3049	3053	3073	3066+101	3053	3073	3053	3073
	7	7	8	8	7	8	9	9	8	6	7	10	9	4	9	5
Homeologs with parental legacy (%)	58.3	58.3	66.7	66.7	58.3	66.7	75.0	69.2	61.5	46.2	53.8	76.9	75.0	33.3	69.2	38.5
	2	3	1	3	2	3	2	2	4	2	4	2	2	4	3	4
	3	2	3	1	3	1	1	2	1	4	2	1	1	4	1	4
Homeologs with expression bias(%)	41.7	41.7	33.3	33.3	41.7	33.3	25.0	30.8	38.5	46.2	46.2	23.1	25.0	66.7	30.8	61.5

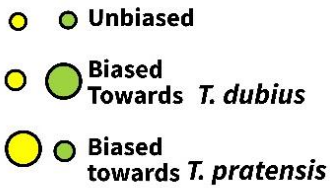


Supplementary Figure 3.8: Comparison of the number of control genes exhibiting homeolog specific expression patterns between polyploids with respect to expression patterns in parent species. Green and yellow circles represent the homeologs derived from *Tragopogon dubius* and *T. pratensis*, respectively. **Parental legacy** (categories a, b & c) is when the expression patterns of genes in the parent species are vertically transmitted in the polyploid species. For example, most of the genes showed no differential expression between parents thus, the homeologs in polyploids showed unbiased expression patterns (category a). **Absence of homeolog expression bias** means the genes were differentially expressed between parent species but in polyploids this trend was absent and homeologs showed no bias (categories d & e). **Novel homeolog expression bias** represents expression bias towards either parent in polyploids although such patterns were not originally observed in either parent species (categories f, g, h & i).

* Early age samples of both late and early flowering plant combined.

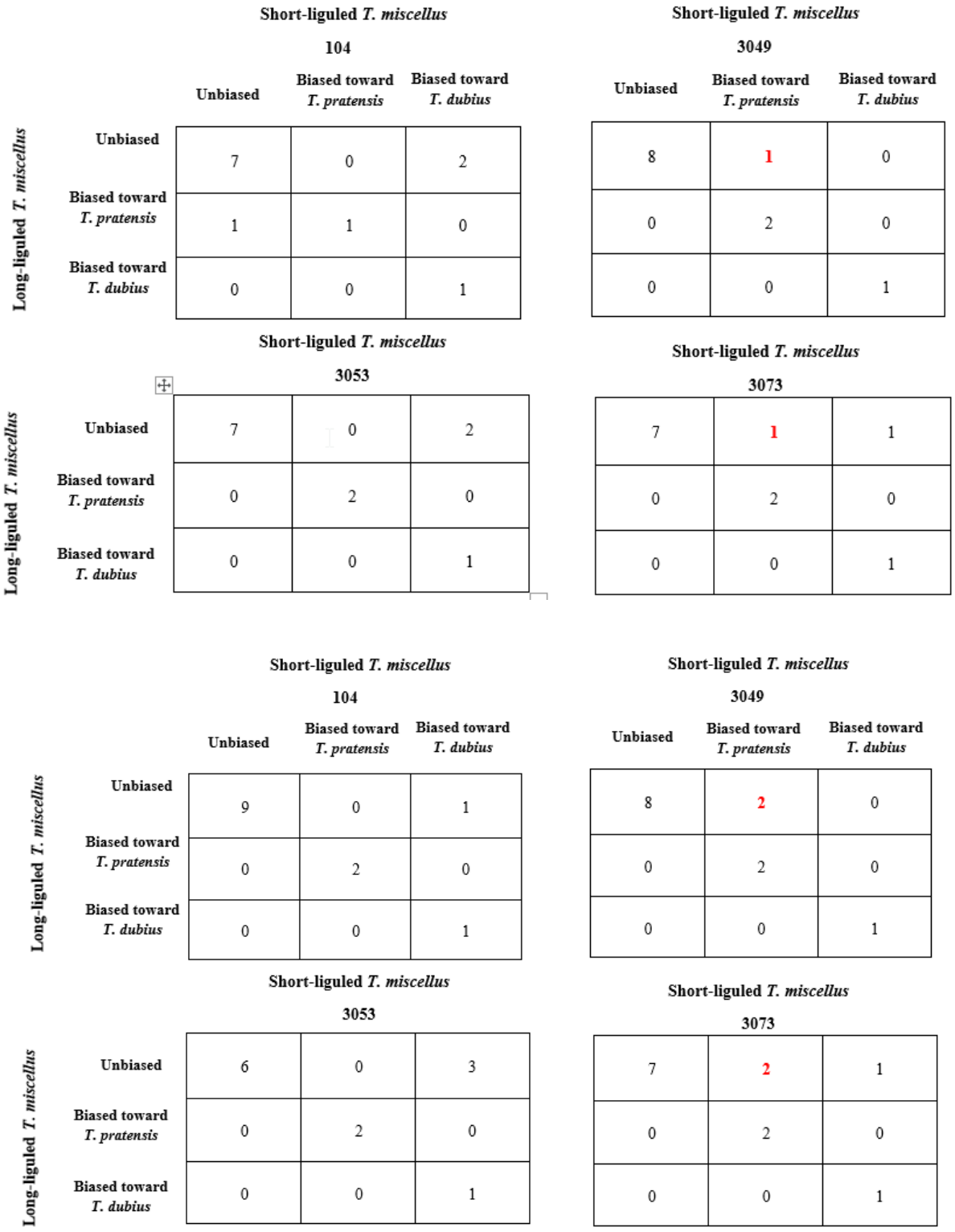


	Late flowering gorup												Early flowering group			
	Early stage		Early stage					Middle stage					Early vs Mature stage			
	Short-liguled *T. miscellus*						Long-liguled *T. miscellus*	Short-liguled *T. miscellus*				Long-liguled *T. miscellus*	Short-liguled *T. miscellus*			
	3053*	3073*	104	3049	3053	3073	3066+101	104	3049	3053	3073	3066+101	3053	3073	3053	3073
	4	5	4	4	4	4	4	4	3	4	4	4	4	3	4	3
Homeologs with parental legacy (%)	57.1	71.4	57.1	57.1	57.1	57.1	57.1	57.1	42.9	57.1	57.1	57.1	57.1	42.9	57.1	42.9
	1		1	1	1	1	2	1	2	2	1	2	1	1	1	1
	2	2	2	2	2	2	1	2	2	1	2	1	2	3	2	3
Homeologs with expression bias(%)	42.9	28.6	42.9	42.9	42.9	42.9	42.9	42.9	57.1	42.9	42.9	42.9	42.9	57.1	42.9	57.1

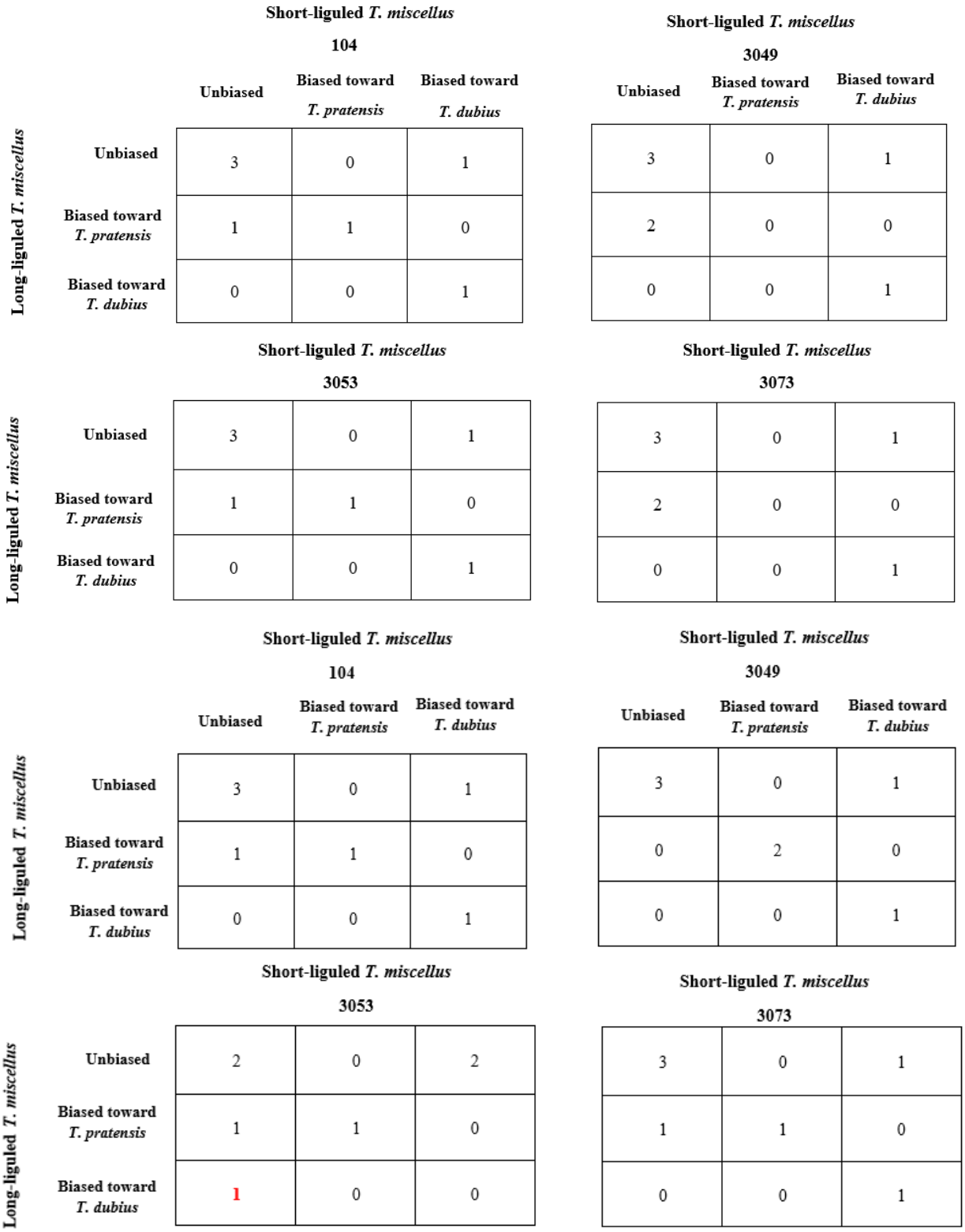


Supplementary Figure 3.9: Comparison of the number of retrograde-signaling partners of dual-targeted genes showing homeolog specific expression patterns between polyploids with respect to expression patterns in parent species. Green and yellow circles represent the homeologs derived from *Tragopogon dubius* and *T. pratensis*, respectively. **Parental legacy** (categories a, b & c) is when the expression patterns of genes in the parent species are vertically transmitted in the polyploid species. For example, most of the genes showed no differential expression between parents thus, the homeologs in polyploids showed unbiased expression patterns (category a). **Absence of homeolog expression bias** means the genes were differentially expressed between parent species but in polyploids this trend was absent and homeologs showed no bias (categories d & e). **Novel homeolog expression bias** represents expression bias towards either parent in polyploids although such patterns were not originally observed in either parent species (categories f, g, h & i).

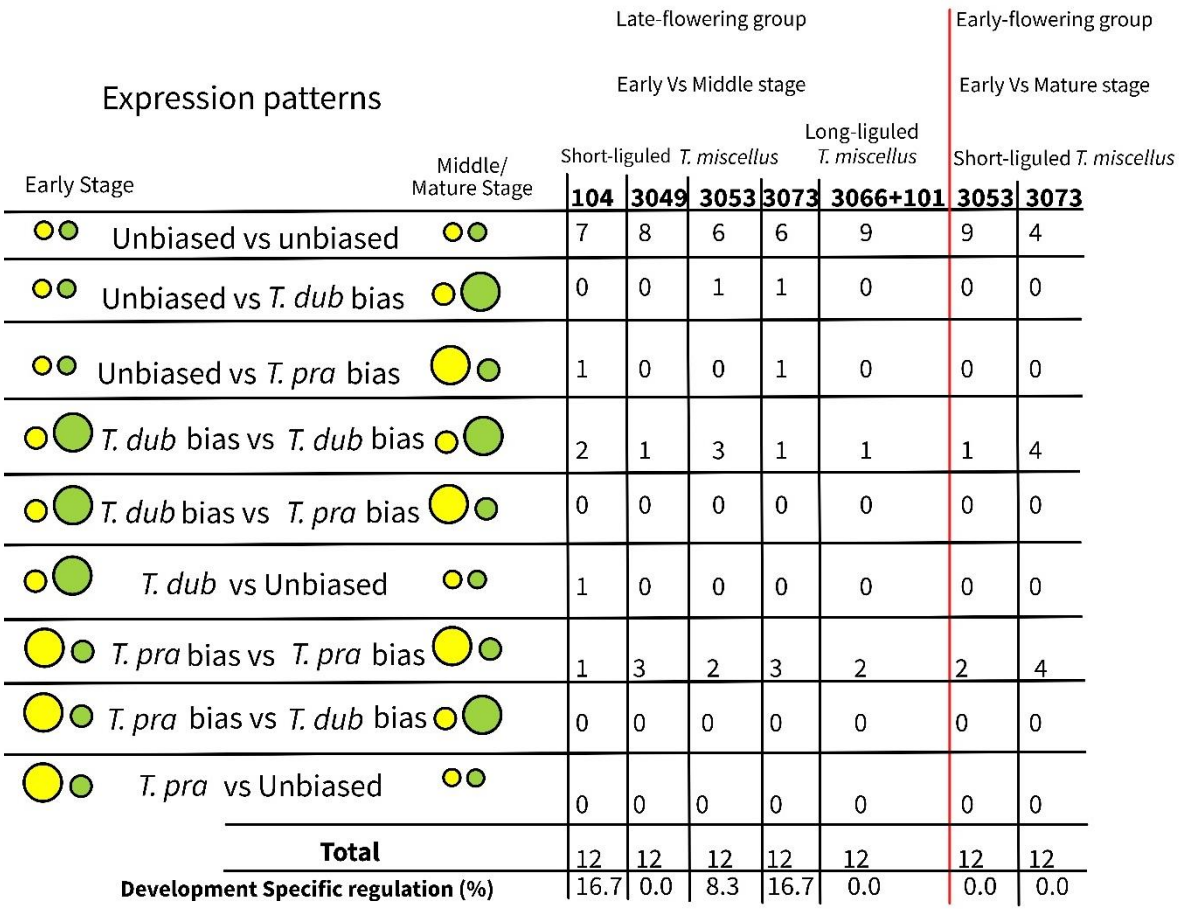
* Early age samples of both late and early flowering plant combined.



Supplementary Figure 3.10: Comparison of the number of control genes belonging to expression patterns including unbiased, bias toward *Tragopogon pratensis* and *T. dubius* between each population of short-liguled *T. miscellus* and long-liguled *T. miscellus* at early- (top) and middle- (bottom) growth stages. The genes in bold red represented the cross specific maternal bias effect.



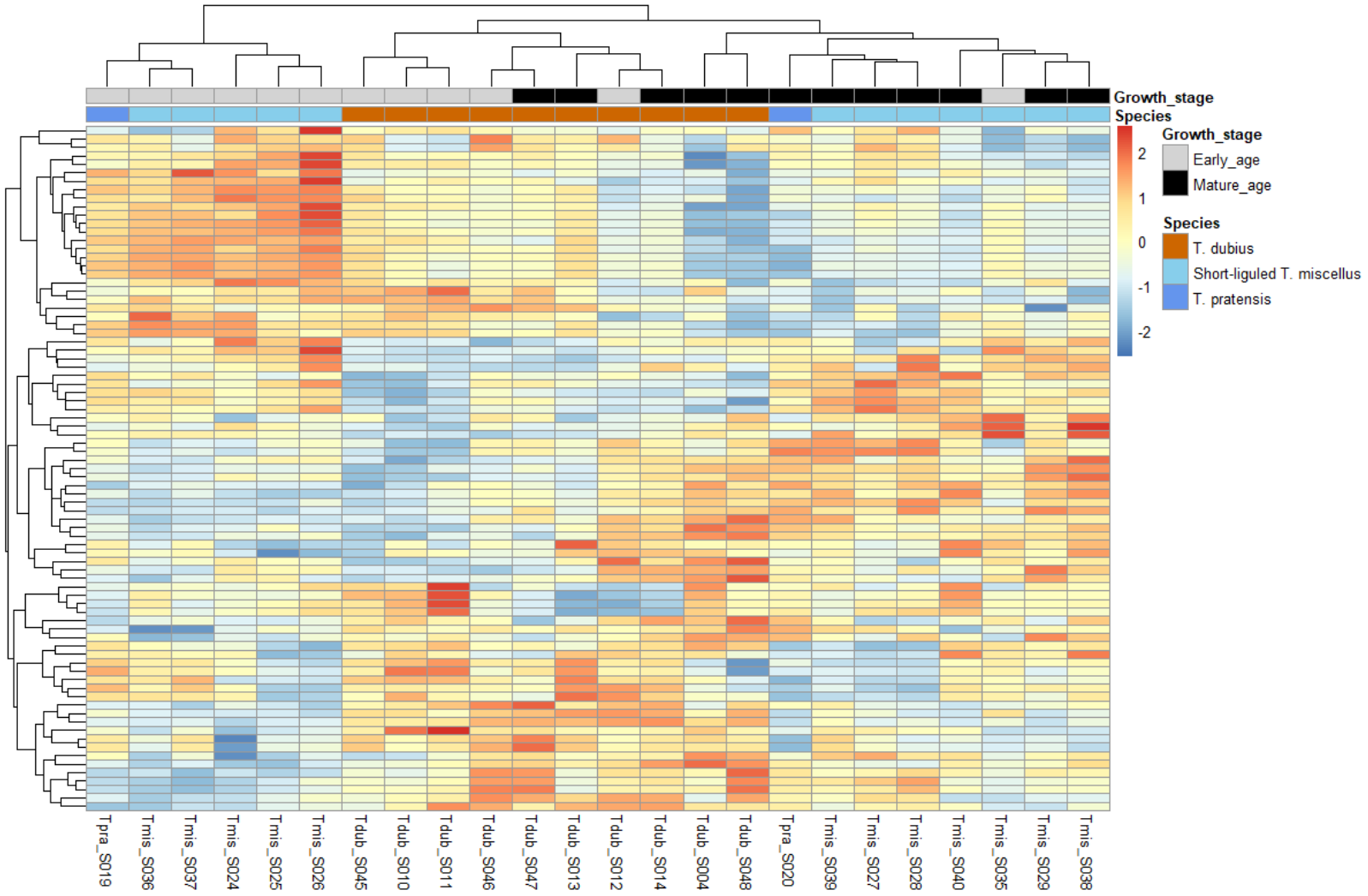
Supplementary Figure 3.11: Comparison of the number of retrograde-signaling genes belonging to expression patterns including unbiased, bias toward *Tragopogon pratensis* and *T. dubius* between each population of short-liguled *T. miscellus* and long-liguled *T. miscellus* at early- (top) and middle- (bottom) growth stages. The genes in bold red represented the cross specific maternal bias effect.



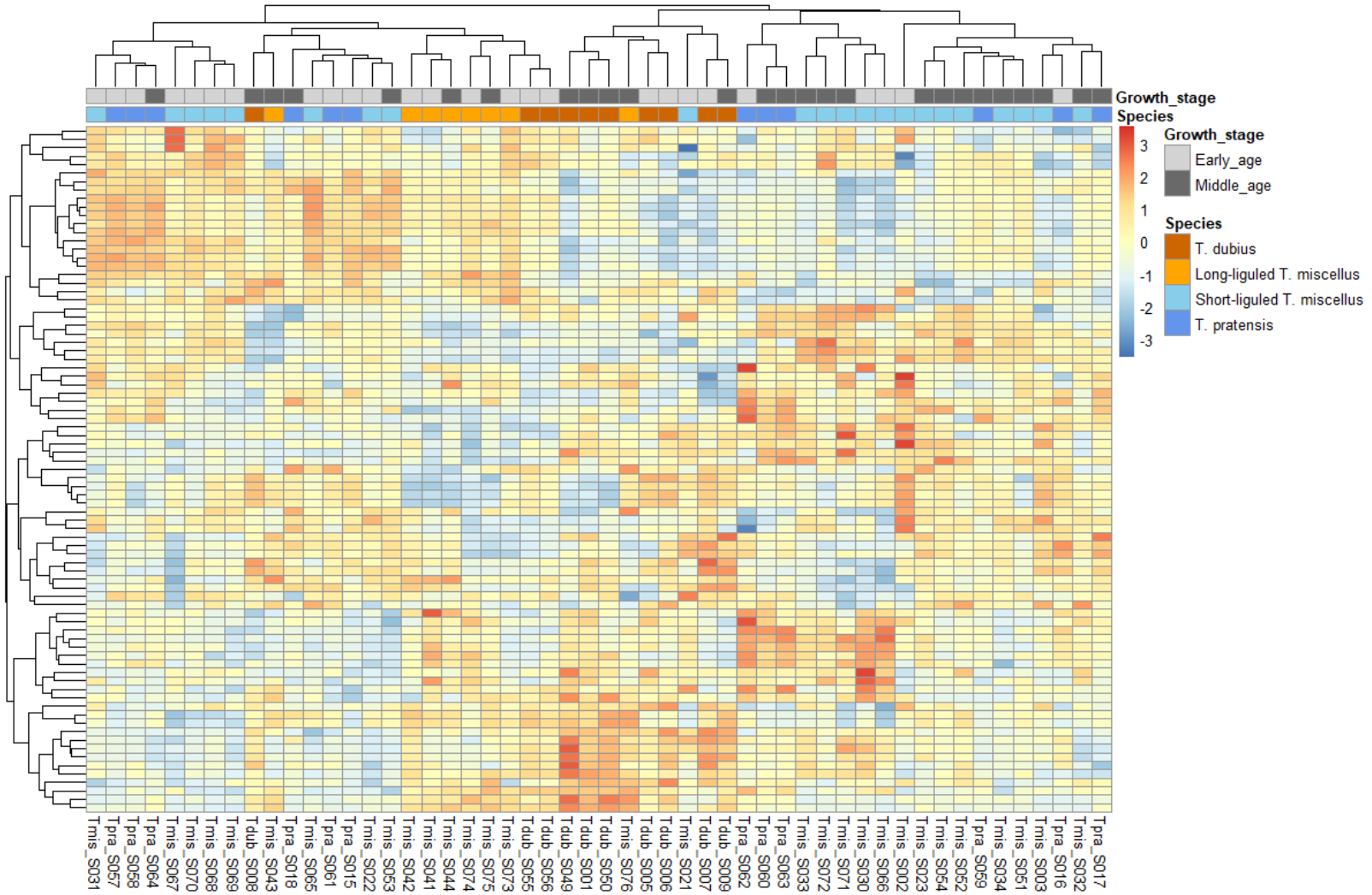
Expression patterns		Late-flowering group					Early-flowering group	
		Early Vs Middle stage					Early Vs Mature stage	
		Short-liguled *T. miscellus*				Long-liguled *T. miscellus*	Short-liguled *T. miscellus*	
Early Stage	Middle/ Mature Stage	**104**	**3049**	**3053**	**3073**	**3066+101**	**3053**	**3073**
Unbiased vs unbiased		7	8	6	6	9	9	4
Unbiased vs *T. dub* bias		0	0	1	1	0	0	0
Unbiased vs *T. pra* bias		1	0	0	1	0	0	0
T. dub bias vs *T. dub* bias		2	1	3	1	1	1	4
T. dub bias vs *T. pra* bias		0	0	0	0	0	0	0
T. dub vs Unbiased		1	0	0	0	0	0	0
T. pra bias vs *T. pra* bias		1	3	2	3	2	2	4
T. pra bias vs *T. dub* bias		0	0	0	0	0	0	0
T. pra vs Unbiased		0	0	0	0	0	0	0
Total		12	12	12	12	12	12	12
Development Specific regulation (%)		16.7	0.0	8.3	16.7	0.0	0.0	0.0

Supplementary Figure 3.12: Comparison of the number of control genes demonstrating homeolog expression patterns between early vs middle growth stage and early vs mature growth stage within each population of short- and long-liguled *Tragopogon miscellus*. Green and yellow circles represent the homeologs derived from *T. dubius* (*T. dub*) and *T. pratensis* (*T. pra*), respectively. The orange highlight across all populations means no directional homeolog bias with respect to a developmental stage.

Supplementary Figure 3.13: Comparison of the number of retrograde-signaling genes displaying homeolog expression patterns between early vs middle growth stage and early vs mature growth stage within each population of short- and long-liguled *Tragopogon miscellus*. Green and yellow circles represent the homeologs derived from *T. dubius* (*T. dub*) and *T. pratensis* (*T. pra*), respectively. The orange highlight across all populations means no directional homeolog bias with respect to a developmental stage.



Supplementary Figure 3.14: Hierarchical clustering of expression data for all organellar (mitochondrial and plastid) genes from individuals of *Tragopogon dubius, T. pratensis* and reciprocal forms of *T. miscellus* in the early-flowering plant group at all growth stages. Differential gene expression is displayed in colors ranging from red to blue as shown in the legend. The positive to negative values on the scale bar correspond to up- and down-regulation of gene expression, respectively.



Supplementary Figure 3.15: Hierarchical clustering of expression data for all organellar (mitochondrial and plastid) genes from individuals of *Tragopogon dubius*, *T. pratensis* and reciprocal forms of *T. miscellus* in the late-flowering plant group at all growth stages. Differential gene expression is displayed in colors ranging from red to blue as shown in the legend. The positive to negative values on the scale bar correspond to up- and down-regulation of gene expression, respectively.

Chapter 4 Supplementary Data

Supplementary Table 4.1: Leaf length measurements (in cm) of individual plant samples as well as average leaf length across each population of diploid parent species; *Tragopogon dubius* and *T. pratensis*, and reciprocal forms of *T. miscellus* allopolyploids. A, B, C and D represent different lineages examined within each population and small letters (a, b, c) show three plant replicates for each individual lineage. Empty cells mean either plants did not germinate or died by the time of measurements (9 weeks post-germination).

		A			B			C			D			Average leaf length (cm)
Species	**Population ID**	**a**	**b**	**c**	**a**	**b**	**c**	**a**	**b**	**c**	**a**	**b**	**c**	
T. dubius	**106**	26.3			27.8	24.8								26.30
	3050	28	10.3	30.5				27.6	28.8	21.6	21.2	29.6		24.70
	3055	27.6	26.5	11.7	26.4	23.6	17.6	25		24.4				22.85
T. pratensis	**3052**					29.6	22	28.8		27.8	25.2		27	26.73
	3058				31.5	31.3	25.5	24.9		23.3	25.4	32.3	24.5	27.34
T. miscellus **(short-liguled)**	**3049**	33.2	32	26.8	36.2	34.4		31.7	35.6	36.7				33.33
	104					27.9	35.7					33.7	34.6	32.98
	3073	33.4	35	38.1		38.3	38	30.4	33.1	35.2	31.9	33.7	33.7	34.62
	3053	38.1	24.5	31.5	32.9	33.8	30.3	32.4	31.9	35.1	35.8	37.2	35.6	33.26
T. miscellus **(long-liguled)**	**101**		33.2	32.4	30.6	30.4	32.2	34.6	34.3	32.9	25.1	29.9		31.56
	3066		30.7	31.5		20.8	30.7		34.3	34.6				30.43

Supplementary Table 4.2: Leaf number of individual plant samples as well as average leaf number across each population of diploid parent species; *Tragopogon dubius* and *T. pratensis*, and reciprocal forms of *T. miscellus* allopolyploids. A, B, C and D represent different lineages examined within each population and small letters (a, b, c) show three plant replicates for each individual lineage. Empty cells mean either plants did not germinate or died by the time of measurements (9 weeks post-germination).

		A			B			C			D			
Species	**Population ID**	**a**	**b**	**c**	**a**	**b**	**c**	**a**	**b**	**c**	**a**	**b**	**c**	**Average leaf number**
T. dubius	**106**	13			22	11								15.33
	3050	14	2	21				18	18	9	7	15		13
	3055	17	19	6	13	12	3	13		8				11.38
T. pratensis	**3052**					7	5	7		7	7		8	6.83
	3058				11	10	7	7		6	6	13	7	8.38
***T. miscellus* (short-liguled)**	**3049**	7	6	6	8	7		7	8	9				7.25
	104					6	7					8	7	7
	3073	8	8	8		8	10	8	6	9	8	9	8	8.18
	3053	9	5	7	9	9	9	7	7	7	7	8	8	7.67
***T. miscellus* (long-liguled)**	**101**		8	7	6	8	7	8	8	8	4	7		7.1
	3066		8	9		4	5		8	8				7

Supplementary Table 4.3: List of germinated individuals from natural populations of diploid parent species, *Tragopogon dubius* and *T. pratensis* as well as reciprocal forms of *T. miscellus* included in this study.

Species	Population	Population ID	Individual ID	Number of seeds germinated out of 3
T. dubius	Pullman, Washington, USA	3040	3	2
T. dubius	Pullman, Washington, USA	3040	10	1
T. dubius	Moscow, Idaho, USA	106	1	1
T. dubius	Moscow, Idaho, USA	106	2	3
T. dubius	Spangle, Washington, USA	3050	1	3
T. dubius	Spangle, Washington, USA	3050	3	3
T. dubius	Spangle, Washington, USA	3050	4	3
T. dubius	Rosalia, Washington, USA	3072	6	1
T. dubius	Oakesdale, Washington, USA	3055	8	3
T. dubius	Oakesdale, Washington, USA	3055	12	3
T. dubius	Oakesdale, Washington, USA	3055	11	2
T. pratensis	Oakesdale, Washington, USA	3052	7	2
T. pratensis	Oakesdale, Washington, USA	3052	17	3
T. pratensis	Oakesdale, Washington, USA	3052	18	2
T. pratensis	Garfield, Washington, USA	3058	9	3
T. pratensis	Garfield, Washington, USA	3058	8	2
T. pratensis	Garfield, Washington, USA	3058	10	3
T. miscellus (long-liguled)	Pullman, Washington, USA	3066	11	2
T. miscellus (long-liguled)	Pullman, Washington, USA	3066	10	2
T. miscellus (long-liguled)	Pullman, Washington, USA	3066	15	2
T. miscellus (long-liguled)	Pullman, Washington, USA	101	4	3
T. miscellus (long-liguled)	Pullman, Washington, USA	101	5	3
T. miscellus (long-liguled)	Pullman, Washington, USA	101	6	3
T. miscellus (long-liguled)	Pullman, Washington, USA	101	10	2
T. miscellus (short-liguled)	Spangle, Washington, USA	3049	1	3
T. miscellus (short-liguled)	Spangle, Washington, USA	3049	2	3
T. miscellus (short-liguled)	Spangle, Washington, USA	3049	3	3
T. miscellus (short-liguled)	Spangle, Washington, USA	3049	4	2
T. miscellus (short-liguled)	Garfield, Washington, USA	104	2	2
T. miscellus (short-liguled)	Garfield, Washington, USA	104	4	2
T. miscellus (short-liguled)	Rosalia, Washington, USA	3073	4	3
T. miscellus (short-liguled)	Rosalia, Washington, USA	3073	5	3
T. miscellus (short-liguled)	Rosalia, Washington, USA	3073	6	3
T. miscellus (short-liguled)	Rosalia, Washington, USA	3073	7	3
T. miscellus (short-liguled)	Oakesdale, Washington, USA	3053	2	3
T. miscellus (short-liguled)	Oakesdale, Washington, USA	3053	8	3
T. miscellus (short-liguled)	Oakesdale, Washington, USA	3053	9	3
T. miscellus (short-liguled)	Oakesdale, Washington, USA	3053	13	3

Supplementary Table 4.4: Tukey's multiple comparison test of germination rates. Comparisons are species-species at a significance level of $p < 0.05$.

Tukey's multiple comparisons test	Mean Differences	95.00% Confidence Interval	Significant?	Adjusted P Value
T. dubius vs. *T. mis* Long	0.8638	-0.7362 to 2.464	No	0.494
T. dubius vs. *T. mis* Short	0.4462	-0.8274 to 1.720	No	0.7954
T. pratensis vs. *T. mis* Long	0.8	-0.9603 to 2.560	No	0.6345
T. pratensis vs. *T. mis* Short	0.3825	-1.088 to 1.852	No	0.9039
T. mis Long vs. *T. mis* Short	-0.4175	-1.888 to 1.052	No	0.879

Supplementary Table 4.5: Tukey's multiple comparison test of leaf number between species (significance level of $p < 0.05$).

Tukey's multiple comparisons test	Mean Differences	95.00% Confidence Interval	Significant?	Adjusted P Value
T. dubius vs. *T. mis* Long	5.622	2.914 to 8.329	Yes	<0.0001
T. dubius vs. *T. mis* Short	5.027	2.753 to 7.301	Yes	<0.0001
T. pratensis vs. *T. mis* Long	0.6518	-2.268 to 3.572	No	0.9361
T. pratensis vs. *T. mis* Short	0.05714	-2.466 to 2.580	No	>0.9999
T. mis Long vs. *T. mis* Short	-0.5946	-3.003 to 1.813	No	0.916

Supplementary Table 4.6: Tukey's multiple comparison test for leaf lengths between species (significance level $p < 0.05$).

Tukey's multiple comparisons test	Mean Differences	95.00% Confidence Interval	Significant?	Adjusted P Value
T. dubius vs. *T. mis* Long	-6.964	-10.47 to -3.461	Yes	<0.0001
T. dubius vs. *T. mis* Short	-9.495	-12.44 to -6.553	Yes	<0.0001
T. pratensis vs. *T. mis* Long	-4.059	-7.837 to -0.2807	Yes	0.0303
T. pratensis vs. *T. mis* Short	-6.59	-9.855 to -3.325	Yes	<0.0001
T. mis Long vs. *T. mis* Short	-2.531	-5.647 to 0.5845	No	0.152