plant biology

Plant Biology ISSN 1435-8603

RESEARCH ARTICLE

Comparative analysis of plastid genomes from allopolyploid *Tragopogon miscellus* and its diploid parents

U. Mukhtar[1], S. C. Newmarch[1], R. C. Winkworth[1], P. S. Soltis[2], D. E. Soltis[2,3] & J. A. Tate[1]

1 School of Agriculture and Environment, Massey University, Palmerston North, New Zealand
2 Florida Museum of Natural History, University of Florida, Gainesville, Florida, USA
3 Department of Biology, University of Florida, Gainesville, Florida, USA

Keywords
Asteraceae; introgression; plastome; polyploidy; *Tragopogon*.

Correspondence
J. A. Tate, School of Agriculture and Environment, Massey University, Palmerston North 4442, New Zealand.
E-mail: j.tate@massey.ac.nz

Editor
X.-.Q. Wang

Received: 13 January 2025;
Accepted: 4 August 2025

doi:10.1111/plb.70109

ABSTRACT

- *Tragopogon* is a model system for the study of recent, recurrent, and reciprocal allopolyploid formation. Recent research has focused on the fates of nuclear genes duplicated in the allopolyploid *T. miscellus* relative to the parental diploids, *T. dubius* and *T. pratensis*. In contrast, little attention has been given to organellar genomes, which interact with the duplicated nuclear genomes via their gene products. Here we reconstructed plastid genomes (plastomes) for representatives of these three species to investigate their structure and variability among natural and synthetic allopolyploids.
- Genomic libraries were Illumina-sequenced for several individuals of the allopolyploid *T. miscellus* and its diploid parents. Whole plastomes were assembled from skimmed data with comparative analyses used to quantify structural and nucleotide variation.
- *Tragopogon* plastomes have a typical quadripartite structure and are similar in size to those of other Asteraceae. The 12 plastomes were highly similar, sharing ~99.5%–100% identity. In all but one case, the plastome sequence for each of the polyploids was most similar to that of its expected maternal parent. The exception involved a polyploid that unexpectedly had a *T. dubius* plastome type, likely as the result of backcrossing with its presumed paternal parent. Such backcrossing events may have contributed to the demise of this polyploid population.
- Plastome sequences can be used to infer the maternal origins of polyploids as well as investigate ongoing population-level dynamics. More fully assessing plastome variation across the geographic distribution of polyploids and their diploid progenitors may provide additional insights into polyploid formation, population dynamics, and subsequent evolution.

INTRODUCTION

Polyploidy, or whole-genome duplication, has been an important process in the evolutionary history of eukaryotes, especially in plants (Vamosi & Dickinson 2006; Jiao *et al.* 2011; Wendel 2015; Soltis *et al.* 2016; Van de Peer *et al.* 2020; Heslop-Harrison *et al.* 2022). Polyploids can be generated within (autopolyploids) or between species (allopolyploids) (Stebbins 1947; Ramsey & Schemske 1998), leading to diverse lineages (Soltis & Soltis 2000; Jiao *et al.* 2011; Wendel 2015; Barker *et al.* 2016). Genetic and genomic studies of polyploids have focused traditionally on the nuclear component, especially the fates of duplicated genes and genomic regions (Chen & Ni 2006; Yoo *et al.* 2013, 2024; Cheng *et al.* 2016; Doyle & Coate 2019). However, plant cells also contain two organellar genomes (plastid and mitochondrial) that remain understudied in polyploids to date (Chen *et al.* 2020; Sloan *et al.* 2024). The dynamics of the organellar genomes in newly formed allopolyploids are of particular interest given that these genomes must interact with additional versions of the nuclear genome (Sloan 2015; Sharbrough *et al.* 2017; Sloan *et al.* 2024). These interactions involve core physiological functions, including photosynthesis and respiration, which are key to the successful establishment of the polyploid lineage (Levin 2003; Rand *et al.* 2004; Pennisi 2016; Sloan *et al.* 2024). Thus, the level of divergence between the organellar genomes of the parental species may offer insights into the compatibility of the nuclear and organellar genomes in the polyploid (Zuo *et al.* 2022; Sloan *et al.* 2024).

In flowering plants, the organellar genomes are typically inherited maternally, although paternal or biparental inheritance has been documented (Corriveau & Coleman 1988; Ruhlman & Jansen 2014). As a result, organellar genome sequences have been used to identify the maternal parent of hybrids and allopolyploids (e.g., Gornicki *et al.* 2014; Chen *et al.* 2020, 2021; Wang *et al.* 2022). Often evolutionary studies have focused on plastomes because these tend to be structurally stable and have variable regions suitable for phylogenetic or population genetic studies (Wicke *et al.* 2011; Ruhlman & Jansen 2014). In contrast, plant mitochondrial genomes are used less often because they tend to be structurally highly variable and exhibit lower rates of nucleotide substitution in plants (Mower *et al.* 2012). Recent efforts to accumulate plastome sequences within species, so-called pan-plastomes, have provided novel insights into species diversity (e.g., Magdy *et al.* 2019; Kan *et al.* 2024; Wang *et al.* 2024), as well as the

origins and evolutionary dynamics of polyploids such as cotton (Yan *et al.* 2024) and wheat (Chen *et al.* 2020). These endeavours get us closer to a fuller understanding of plant pangenomes across polyploids (Shan *et al.* 2025), where studies of non-crop polyploids are still lacking.

Three *Tragopogon* diploids (*T. dubius* Scop., *T. pratensis* L., and *T. porrifolius* L.) and their derivative allopolyploids (*T. mirus* Ownbey and *T. miscellus* Ownbey) are a model system for investigating the early stages of genome evolution following polyploidy (Soltis *et al.* 2004, 2012). The two allopolyploid species formed in the early 20th century following the introduction of the diploids into the northwestern United States from Europe and Asia (Ownbey 1950; Soltis *et al.* 2022, 2023). *Tragopogon mirus* formed from *T. dubius* and *T. porrifolius*, while *T. miscellus* formed from *T. dubius* and *T. pratensis* (Ownbey 1950; Ownbey & McCollum 1953). In nature, *T. miscellus* formed reciprocally from its diploid progenitors (Ownbey 1950; Ownbey & McCollum 1953). The reciprocal forms of *T. miscellus* differ strikingly in their inflorescence and floral morphology. When *T. dubius* is the maternal parent, the capitulum is more open and the florets have longer ligules (fused petals), and when *T. pratensis* is the maternal parent, the capitulum is constricted at the base and the florets have shorter ligules (Fig. 1; Ownbey & McCollum 1953; Soltis & Soltis 1989; Tate *et al.* 2009). These parent-of-origin morphologies are also seen in synthetic allopolyploids (Tate *et al.* 2009). In contrast to *T. miscellus*, *T. mirus* is only known to have formed in nature when *T. porrifolius* acted as the maternal parent (Ownbey 1950; Ownbey & McCollum 1953).

As is common for most polyploids (Ashton & Abbott 1992; Brochmann *et al.* 1992a, 1992b; Segraves *et al.* 1999; Soltis & Soltis 1999; Doyle *et al.* 2002), both *Tragopogon* allopolyploids have formed recurrently. Ownbey (1950) originally described the tetraploid populations as 'small and precarious,' and since then, several populations of both diploids and allopolyploids have gone extinct (Novak *et al.* 1991; Soltis *et al.* 2004), with new polyploid populations also identified (e.g., Soltis *et al.* 2012). Consistent with the observation of recurrent formation, molecular analyses have shown that independently formed allopolyploid populations are genetically distinct (Soltis & Soltis 1989; Cook *et al.* 1998; Symonds *et al.* 2010).

Several previous studies investigated nuclear genome evolution in the *Tragopogon* allotetraploids and found that often one of the duplicated nuclear gene copies (homeologs) was lost, and that gene expression of the parental homeologs differed among individuals both within and between populations (Tate *et al.* 2006; Buggs *et al.* 2009, 2011, 2012; Koh *et al.* 2010; Sehrish *et al.* 2015; Shan *et al.* 2020). A study of restriction fragment length polymorphisms (RFLPs) of plastid DNA found differences among populations of the diploid *Tragopogon* species and that this variation was shared with different allopolyploid populations (Soltis & Soltis 1989). Notably, these RFLPs allowed the maternal progenitor of independently formed allopolyploid populations to be identified (Soltis & Soltis 1989). Despite this insight, subsequent studies have not utilised plastid DNA sequences to examine the origins and evolution of *Tragopogon* allopolyploids. An investigation of plastome structure and sequence divergence in *Tragopogon* allopolyploids and their diploid progenitors would provide valuable insights into both their origins and patterns of cytonuclear coordination.

Here we report annotated plastid genome sequences for representatives of *Tragopogon miscellus* and its diploid parents. We used these data to compare plastome structure and levels of variation between (1) the diploid parental species, *T. dubius* and *T. pratensis*, (2) the parental species and their allopolyploid derivative, *T. miscellus*, (3) reciprocally formed natural and synthetic *T. miscellus* individuals, and (4) independently formed short-liguled *T. miscellus* populations. Additionally, for the reciprocally formed allopolyploids, we aimed to assess

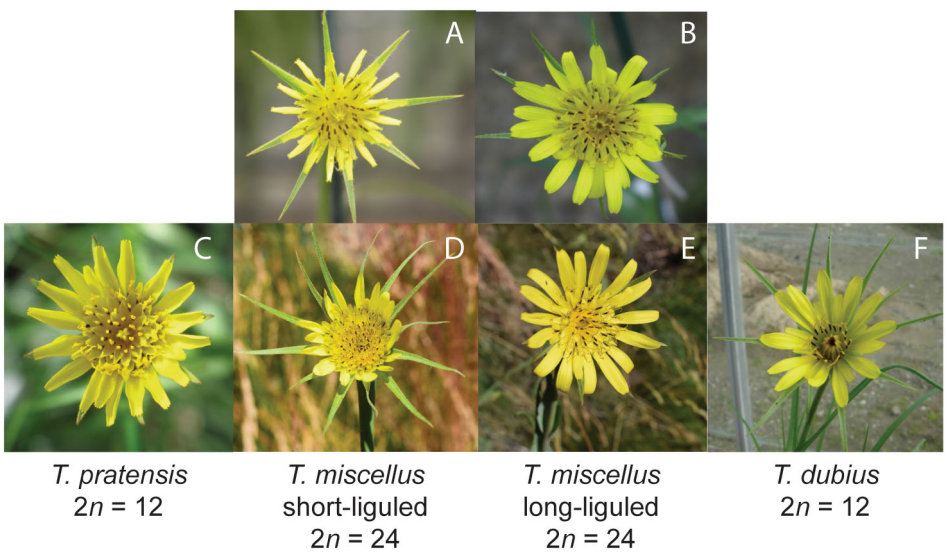

Fig. 1. Capitula of reciprocally formed *Tragopogon miscellus* allopolyploids and diploid parents, *T. pratensis* and *T. dubius*. Synthetic short-liguled *T. miscellus* (A) has *T. pratensis* (C) as its maternal parent and *T. dubius* (F) as its paternal parent. Synthetic long-liguled *T. miscellus* (B) is formed reciprocally with *T. dubius* (F) as the maternal parent and *T. pratensis* as paternal parent (C). Naturally formed short-liguled (D) and long-liguled (E) *T. miscellus* have the same parental combinations as the synthetics.

Table 1. Details of the sequenced *Tragopogon* samples including their ploidal level and form (for *T. miscellus*), collection locations, and lineage identifiers (corresponding to Soltis and Soltis collection numbers or cross numbers from Tate *et al.* 2009).

species	ploidal level/form	population	lineage
Tragopogon dubius	2*x*	Pullman, Washington	2613-11
Tragopogon dubius	2*x*	Pullman, Washington	2613-24
Tragopogon dubius	2*x*	Spangle, Washington	2616-1
Tragopogon pratensis	2*x*	Moscow, Idaho	2608-35
Tragopogon miscellus	4*x*/short-liguled	Moscow, Idaho	2604-4
Tragopogon miscellus	4*x*/short-liguled	Albion, Washington	2625-3
Tragopogon miscellus	4*x*/short-liguled	Albion, Washington	2625-6
Tragopogon miscellus	4*x*/short-liguled	Spokane, Washington	2664-5
Tragopogon miscellus	4*x*/short-liguled	Garfield, Washington	2688-9
Tragopogon miscellus	4*x*/short-liguled	Synthetic	111-7-5
Tragopogon miscellus	4*x*/long-liguled	Pullman, Washington	2605-23
Tragopogon miscellus	4*x*/long-liguled	Synthetic	129-7-10

whether the maternal origin of the plastome sequences could be identified.

MATERIALS AND METHODS

Taxon sampling, DNA extraction, and Illumina short-read sequencing

Tissue samples were sourced from 12 existing collections (Tate *et al.* 2009; Sehrish *et al.* 2014, 2015); details are provided in Table 1. The diploid parents of *T. miscellus* were represented by three samples from two populations of *Tragopogon dubius* and one of *T. pratensis*. Sampling of the short-liguled form of *T. miscellus* included five individuals from four natural populations, as well as a second-generation synthetic individual (Tate *et al.* 2009). The long-liguled form of *T. miscellus* was represented by an individual from one natural population of this form (Pullman, WA, USA) and a second-generation synthetic long-liguled individual (Tate *et al.* 2009).

Total DNA was extracted using a modified CTAB method (Doyle & Doyle 1987) or the Qiagen DNeasy Plant Mini Kit (Qiagen, Germantown, USA) following the manufacturer's instructions. The quality of extracted DNA was assessed by 1.5% agarose gel electrophoresis. DNA quantity was determined using a Qubit 2.0 Fluorometer (ThermoFisher Scientific, Waltham, MA, USA) and Qubit dsDNA Broad Range Assay kit (ThermoFisher Scientific) following the manufacturer's instructions. Massey Genome Service (Palmerston North, New Zealand) prepared sequencing libraries from total DNA using Illumina Nextera DNA library preparation kits (Illumina, Inc., San Diego, CA, USA). The libraries were pooled, and paired-end sequencing (250 or 300 bp) was performed on Illumina MiSeq or HiSeq instruments.

Plastome assembly and annotation

For each sample, a complete plastome sequence was assembled from quality-assessed paired-end Illumina reads using a combination of de novo and reference-based approaches (Straub *et al.* 2012; Twyford & Ness 2017). Sequence reads were first assembled into contigs using idba_ud v. 1.1.3 (Peng *et al.* 2012). The resulting contigs were then compared to the plastome sequence of *Lactuca sativa* L. (GenBank accession NC_007578) using BLAST v. 2.3 (Altschul *et al.* 1997). Contigs with high similarity to the reference were assembled into larger scaffolds and ultimately draft plastome sequences using the de novo and reference-based assembly tools implemented in Geneious R9 (Kearse *et al.* 2012). To produce a circularised draft plastome, scaffolds that contained genes typically located at the inverted repeat (IR) boundaries (e.g., *rps-22*, *ycf1*) were identified. These sequences typically contained two distinct portions (i.e., a segment of the IR and one or other end of a single-copy region), and by comparing across them, it was possible to identify the IR boundaries. The IR region was then manually copied, inverted, and incorporated into the draft plastome sequence. Draft plastome sequences were checked for assembly errors by mapping the quality-assessed reads to drafts using BWA v. 0.7.17 (Li & Durbin 2009). Draft sequences were edited as appropriate following a visual inspection of mapped reads in Tablet v. 11.0.10 (Milne *et al.* 2009). We also used GetOrganelle (Jin *et al.* 2020) as a secondary assembly method using *T. dubius* (2613-24) as the reference, but as there were no appreciable differences in the results, those data are not presented here.

A draft annotation of the *Tragopogon dubius* plastome (using individual 2613-24) was prepared by comparison to the annotation of the *Lactuca sativa* reference (GenBank accession NC_007578) and inspected manually. Specifically, protein-coding annotations were checked for appropriate start and stop codons, maintenance of the coding frame, and similarity to existing sequences; RNA annotations that shared >80% identity with corresponding loci in the reference sequences were removed, with the start and end positions of the remaining loci checked for consistency with existing sequences. Annotations for the remaining *Tragopogon* plastomes used the *T. dubius* plastome as the reference.

Data generated as part of this study were deposited in NCBI under accession numbers: PQ572734-PQ572745 (Table 2).

Multiple sequence alignment and comparative analyses

For initial comparisons, we selected publicly available complete plastome sequences from ten representatives of Asteraceae (Table 2). Where necessary, inverted repeats (IRs), large single-copy (LSC), and small single-copy (SSC) regions were identified using Ge-Seq (Tillich *et al.* 2017) and Chloe (https://github.com/ian-small/chloe). For each of these sequences, the overall GC content, as well as that of the IRs, large, and small single-copy regions were estimated using Geneious R9.

To evaluate plastome diversity across the 12 included individuals, complete plastome sequences were aligned using MAFFT (Katoh *et al.* 2002). This alignment was visually examined in Geneious R9 and corrected for minor improvement (e.g., to maintain reading frame). Two short sections of the

Table 2. Plastome sizes and characteristics for newly sequenced *Tragopogon* individuals compared to different tribes and subtribes in the Asteraceae.

species (*Tragopogon* lineage or common name)	tribe/subtribe	sequence length in bp (%GC in parentheses)				GenBank ID
		overall	LSC	SSC	IRs	
Tragopogon dubius (2613-11)	Cichorieae/Scorzonerinae	153,001 (37.7%)	84,225 (35.8%)	18,407 (31.1%)	25,184 (43.1%)	PQ572734
Tragopogon dubius (2613-24)	Cichorieae/Scorzonerinae	153,001 (37.7%)	84,225 (35.8%)	18,407 (31.1%)	25,184 (43.1%)	PQ572735
Tragopogon dubius (2616-1)	Cichorieae/Scorzonerinae	152,982 (37.7%)	84,197 (35.8%)	18,437 (31.1%)	25,173 (43.1%)	PQ572736
Tragopogon miscellus (2604-4)	Cichorieae/Scorzonerinae	153,002 (37.7%)	84,324 (35.8%)	18,332 (31.1%)	25,173 (43.1%)	PQ572740
Tragopogon miscellus (2625-3)	Cichorieae/Scorzonerinae	152,953 (37.7%)	84,168 (35.9%)	18,436 (31.1%)	25,173 (43.1%)	PQ572741
Tragopogon miscellus (2625-6)	Cichorieae/Scorzonerinae	153,009 (37.7%)	84,331 (35.8%)	18,332 (31.1%)	25,173 (43.1%)	PQ572742
Tragopogon miscellus (2664-5)	Cichorieae/Scorzonerinae	153,000 (37.7%)	84,324 (35.8%)	18,330 (31.1%)	25,173 (43.1%)	PQ572743
Tragopogon miscellus (2688-9)	Cichorieae/Scorzonerinae	153,016 (37.7%)	84,324 (35.8%)	18,346 (31.1%)	25,173 (43.1%)	PQ572744
Tragopogon miscellus (111-7-5)	Cichorieae/Scorzonerinae	153,002 (37.7%)	84,324 (35.8%)	18,332 (31.1%)	25,173 (43.1%)	PQ572739
Tragopogon miscellus (2605-23)	Cichorieae/Scorzonerinae	153,001 (37.7%)	84,225 (35.8%)	18,407 (31.1%)	25,184 (43.1%)	PQ572738
Tragopogon miscellus (129-7-10)	Cichorieae/Scorzonerinae	153,001 (37.7%)	84,225 (35.8%)	18,407 (31.1%)	25,184 (43.1%)	PQ572737
Tragopogon pratensis (2608-35)	Cichorieae/Scorzonerinae	153,002 (37.7%)	84,324 (35.8%)	18,332 (31.1%)	25,173 (43.1%)	PQ572745
Scorzonera radiata	Cichorieae/Scorzonerinae	152,631 (37.7%)	83,967 (35.9%)	18,562 (31.1%)	25,051 (43.2%)	NC_087968
Cynara cardunculus (Artichoke)	Cardueae/Carduinae	152,529 (37.7%)	83,578 (35.8%)	18,641 (31.5%)	25,155 (43.2%)	KP842708
Cichorium intybus (Chicory)	Cichorieae/Cichoriinae	152,975 (37.7%)	84,232 (35.8%)	18,562 (31.2%)	25,091 (43.2%)	NC_043842
Taraxacum officinale (Dandelion)	Cichorieae /Crepidinae	151,323 (37.7%)	83,899 (35.9%)	18,562 (31.4%)	24,431 (42.9%)	ON641329
Lactuca sativa (Lettuce)	Cichorieae/Lactucinae	152,765 (37.5%)	84,103 (35.7%)	18,596 (31.1%)	25,033 (43.0%)	NC_007578
Leucanthemum vulgare (Oxeye daisy)	Anthemideae/ Leucantheminae	150,191 (37.5%)	82,675 (35.6%)	18,400 (30.6%)	24,558 (43.0%)	NC_047460
Artemisia annua (Sweet wormwood)	Anthemideae/Artemisiinae	150,952 (37.5%)	82,772 (35.6%)	18,268 (31.0%)	24,956 (43.0%)	NC_034683
Aster tataricus (Tatarian aster)	Astereae/Asterinae	152,992 (37.3%)	84,698 (35.1%)	18,250 (31.2%)	25,022 (43.0%)	NC_042913
Dahlia pinnata (Dahlia)	Coreopsideae	152,107 (37.5%)	83,704 (35.6%)	18,347 (31.2%)	25,028 (43.0%)	NC_066129
Helianthus annuus (Sunflower)	Heliantheae	151,104 (37.6%)	83,530 (35.7%)	18,308 (31.4%)	24,633 (43.2%)	NC_007977

Total plastome size and lengths of the Large Single-Copy (LSC), Small Single-Copy (SSC), and Inverted Repeat (IR) regions are specified, with GC content of each in parentheses. GenBank IDs are also provided.

alignment (i.e., 1–13 nucleotides depending on the individual) were removed where pairs of homopolymer repeats were in tandem and resulted in alignment ambiguity. Variant identification was performed on the full data set in Geneious R9 using the inbuilt variant (polymorphism) calling tool and *T. dubius* (2613-24) as the reference. Variant calling on subsets of the data used the same tool and an appropriate reference. The GC content statistics were again estimated using Geneious R9.

To further assess relationships among the included *Tragopogon* individuals, we conducted a pair of phylogenetic analyses. One analysis used an alignment trimmed to include the LSC, SSC, and one copy of the IR, and the other analysis used an alignment comprising only the coding sequences from the same three regions. To these alignments, we added publicly available sequences from a Chinese representative of *T. dubius* (GenBank accession OR840963), a *T. pratensis* from the United Kingdom (OZ060766), and *Scorzonera radiata* Fisch. ex DC. (Cichorieae/Scorzonerinae; NC_087968) as an outgroup (Mavrodiev *et al.* 2012). Alignment positions were removed when less than 60% of the sequences were represented (e.g., due to length differences) or the alignment was otherwise ambiguous (e.g., due to gene sequence variation). We identified the best-fit partition scheme for each alignment and the best-fit substitution model for each partition using ModelFinder (Kalyaanamoorthy *et al.* 2017), as implemented in IQ-TREE v2.2.0 (Minh *et al.* 2020). Partitioned maximum likelihood searches were then conducted using IQ-TREE v2.2.0 (Chernomor *et al.* 2016; Minh *et al.* 2020). For these searches, the best-fit substitution models were applied to each data partition and support for inferred relationships evaluated using 1,000 bootstrap replicates.

RESULTS

For each of the 12 sampled *Tragopogon* individuals, whole plastomes were assembled from paired-end Illumina short-read sequences (Table 2). Read coverage for these plastome sequences ranged from 111× to 546×, with an average of 350×. The following sections describe various comparisons of the recovered plastomes.

Diploid *Tragopogon* plastomes

The assembled plastome sequences for *Tragopogon dubius* and *T. pratensis* have the quadripartite structure typical of most angiosperms and are consistent in both size and GC content with other representatives of Asteraceae (Fig. 2). For example, plastome sizes were 150,191–152,992 bp for the other 10 representatives of Asteraceae and 152,982–153,002 bp for the *Tragopogon* diploids assembled in the present study (Table 2).

The LSC, SSC, and each IR of the plastome sequences from the diploid *Tragopogon* species were 84,197–84,324 bp, 18,332–18,437 bp, and 25,173–25,184 bp in length, respectively (Table 2). For these plastomes, the LSC was marginally larger and the SSC marginally smaller in *T. pratensis* than in

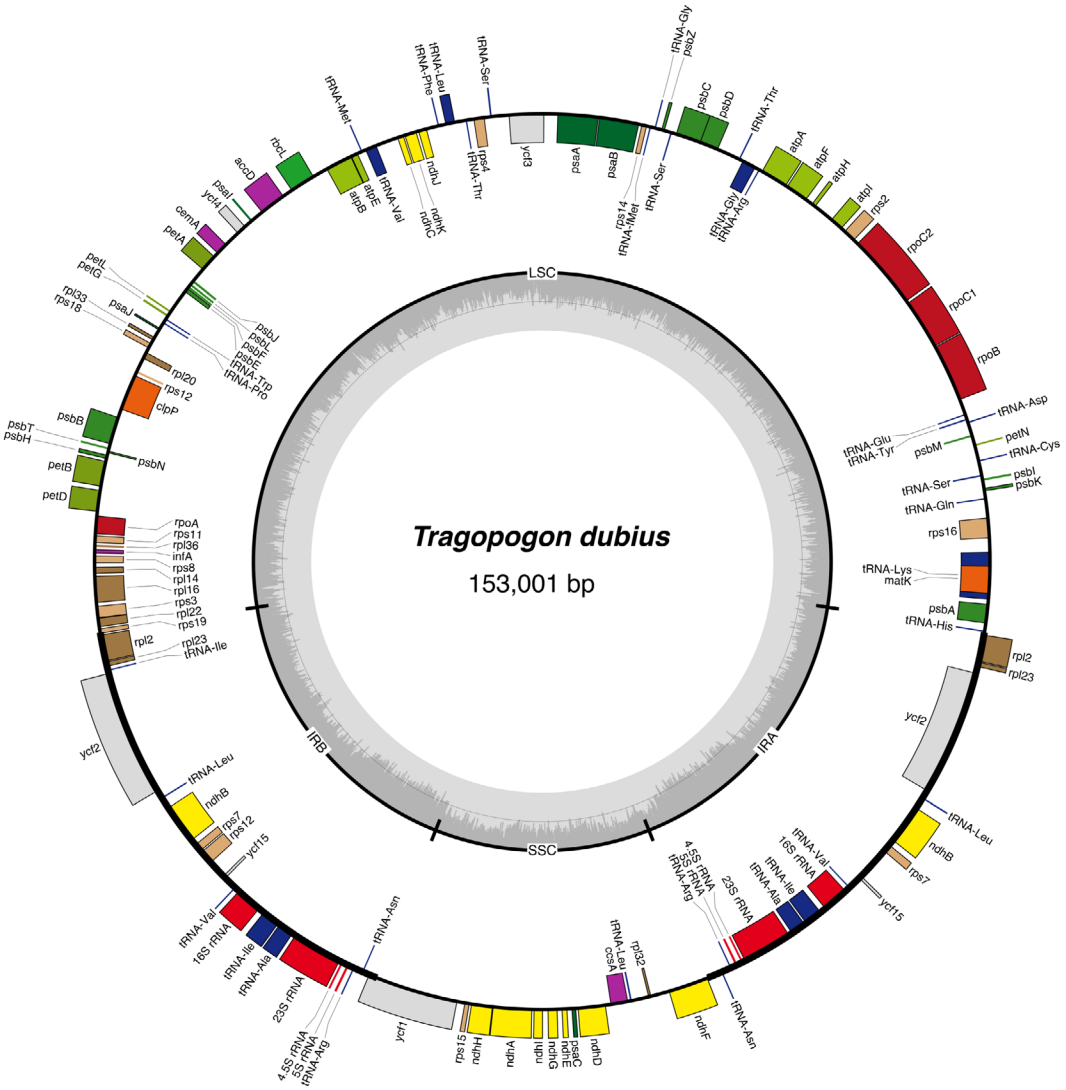

Fig. 2. The plastome of *Tragopogon dubius*. Genes on the interior of the outer circle are transcribed in the forward direction; those on the exterior are transcribed in the reverse direction. The thickened sections on the outer circle indicate the IR regions. The inner circle indicates the inverted repeat (IR_A and IR_B), the large single-copy (LSC) region, and the small single-copy (SSC) region. The shaded ring on the interior of this circle indicates changes in %GC. Plastome drawn using OrganellarGenomeDraw (Lohse *et al*. 2013).

the three representatives of *T. dubius*. Gene annotations for all four diploid individuals were co-linear and comprised 120 gene loci. The gene complement included 83 protein-coding genes, eight rRNAs, and 37 tRNAs that accounted for all 20 amino acids. Based on sequence integrity (i.e., no stop codons were detected), each locus was presumed to be functional.

The plastome sequences of the two *T. dubius* individuals from Pullman (2613-11 and 2613-24) shared 100% identity (Table 3, Table S1). In contrast, the *T. dubius* plastome from Spangle (2616) differed from the Pullman plastomes by 84 polymorphisms (99.8% identity), consisting of 48 single nucleotide polymorphisms (SNPs) and 36 indels, 24 of which involved homopolymer repeats (Tables 3 and 4). Of the 48 SNPs, 14 were in coding sequences, and of these, seven were nonsynonymous substitutions that fell within either the *ycf1* (three substitutions) or *ycf2* (two substitutions per copy) loci (Tables 3 and 4). These genes also contained in-frame indels: one in *ycf1* and one in each of the *ycf2* gene copies (Table 3).

The two *T. dubius* individuals from Pullman (2613-11 and 2613-24) and *T. pratensis* (2608-35) exhibited 99.58% identity, differing by 194 polymorphisms. These differences included 132 substitutions, of which 63 substitutions occurred in coding sequences and 26 involved nonsynonymous changes (Tables 3 and 4). Most nonsynonymous substitutions fell in the *ycf1* (nine substitutions) or *ycf2* (five substitutions per copy) gene sequences (Table 4). Of the remaining seven nonsynonymous substitutions, two occurred in the *matK* (maturase K) gene and one each in the *cemA* (envelope membrane protein), *atpI* (ATP synthase CF0 A subunit), *ndhF* (NADH dehydrogenase F), *rpoB* (RNA polymerase beta subunit), and *rpoC2* (RNA polymerase beta subunit) genes (Table 4). There were also 62 indels, of which 35 involved homopolymer repeats (Table 3).

Table 3. Summary of sequence polymorphisms among the 12 *Tragopogon* plastomes.

	sequence polymorphisms						
		length polymorphisms			nucleotide polymorphisms		
						synonymous/ nonsynonymous substitutions	
group compared	total	overall/within protein-coding sequences	homopolymer repeats	indels	overall/within protein-coding sequences	overall	within *ycf1* and *ycf2*
All 12 samples	241	77/5	41/1	36/4	164/67	38/29	14/22
T. dubius and *T. pratensis*	238	77/5	41/1	36/4	161/69	38/29	14/22
Within *T. dubius*	88	36/3	22/1	14/2	52/15	8/7	7/7
T. dubius (except 2616-1) and long-liguled *T. miscellus*	0	n/a	n/a	n/a	0/0	n/a	n/a
T. pratensis and short-liguled *T. miscellus* (except 2625-3)	0	n/a	n/a	n/a	0/0	n/a	n/a
T. dubius 2616-1 and *T. miscellus* 2625-3	11	6/0	5/0	1/0	5/1	1/0	0/0
T. pratensis and *T. miscellus* 2625-3	214	70/4	36/0	34/4	141/58	33/25	11/18

Table 4. Nonsynonymous substitutions among the 12 *Tragopogon* plastomes.

gene	nucleotide substitution	codon position	amino acid change
ATP synthase CF0 A subunit (*atpI*)	T ↔ C	8	Ile ↔ Thr*
Envelope membrane protein (*cemA*)	C ↔ A	37	Thr ↔ Asn
Maturase K (*matK*)	T ↔ C	4	Phe ↔ Ser*
	G ↔ A	261	Arg ↔ Lys
NADH dehydrogenase F (*ndhF*)	A ↔ C	419	Glu ↔ Ala*
RNA polymerase beta subunit (*rpoB*)	G ↔ A	90	Glu ↔ Lys
RNA polymerase beta subunit (*rpoC2*)	T ↔ G	918	Asn ↔ Lys
Hypothetical chloroplast RF1 (*ycf1*)	G ↔ A	237	Val ↔ Ile
	A ↔ G	262	Asn ↔ Asp
	A ↔ C	468	Lys ↔ Gln
	T ↔ A	514	Phe ↔ Leu
	C ↔ T	540	Pro ↔ Ser*
	A ↔ C	617	Glu ↔ Asp
	A ↔ T	862	Leu ↔ Phe
	T ↔ A	870	Phe ↔ Leu
	C ↔ G	909	Pro ↔ Ala
	A ↔ G	952	Lys ↔ Glu
Hypothetical chloroplast RF1 (*ycf2*)	G ↔ T	348	Glu ↔ Asp
	G ↔ T	1266	Arg ↔ Ile*
	T ↔ G	1273	Ile ↔ Arg*
	A ↔ G	1488	Asn ↔ Ser
	C ↔ G	1863	Arg ↔ Gly*
	G ↔ T	2070	Arg ↔ Ile*

An asterisk indicates that the nucleotide substitution results in a change in amino acid polarity.

Although most indels fell within non-coding spacer regions, the Pullman *T. dubius* and *T. pratensis* plastome sequences were distinguished by the same three indels in the *ycf1* and *ycf2* genes as distinguished the Pullman and Spangle *T. dubius* plastome sequences.

The *T. dubius* individual from Spangle (2616-1) and *T. pratensis* (2608-35) shared 99.51% identity and differed by 203 polymorphisms. In this case, the polymorphisms included 133 SNPs (34 synonymous and 24 nonsynonymous) and 70 indels, of which 35 involved homopolymer repeats.

Comparisons of the diploid and polyploid plastomes

Plastome sequences from the sampled *Tragopogon miscellus* allopolyploids all shared high sequence identity (99.98%–100%) with one or more of the diploid plastome sequences. A pairwise similarity matrix is provided in Table S1.

Diploid *T. dubius* is the maternal parent of the long-liguled form of *T. miscellus*. As expected, plastome sequences from both long-liguled individuals (129-7-10 and 2605-23) shared the highest identity with those from *T. dubius*. More specifically, sequences from both the natural and synthetic long-liguled *T. miscellus* individuals shared 100% identity with those from the two *T. dubius* individuals from Pullman (2613-11 and 2613-24). The long-liguled *T. miscellus* sequences differed from the Spangle *T. dubius* (2616-1) sequence by the same 84 polymorphisms as described in the previous section for the Pullman and Spangle *T. dubius* sequences.

Diploid *T. pratensis* is the maternal parent of the short-liguled form of *T. miscellus*. In this case, five of the six short-liguled *T. miscellus* sequences shared 100% identity with the *T. pratensis* plastome sequence (Table 3). The exception was *T. miscellus* 2625-3 from Albion, which shared 99.52% identity with the *T. pratensis* plastome sequence. These two sequences differed by 206 polymorphisms: 136 substitutions including 33 synonymous and 24 non-synonymous SNPs, plus 70 indels, of which 36 involved homopolymer repeats (Table 3). Unlike the remaining short-liguled *T. miscellus* accessions, the plastome sequence of *T. miscellus* (2625-3) shared higher identity with *T. dubius* sequences (99.83%–99.98%), and, in particular, the Spangle *T. dubius* (2616-1) with which it shared 99.98% identity and differed by 11 polymorphisms (Table 3).

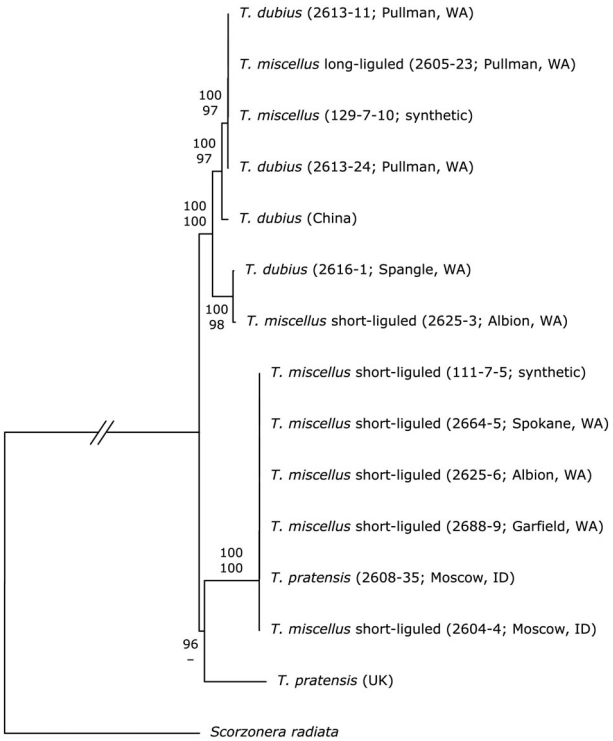

Fig. 3. Maximum likelihood phylogeny for the included accessions of *Tragopogon dubius*, *T. miscellus*, and *T. pratensis*. Numbers associated with branches are bootstrap support values recovered from analyses of the data set containing a single copy of the IR (upper value) and coding sequences (lower value). Support values less than 70% are not shown but are indicated by a dash (–). The branch leading to the outgroup, *Scorzonera radiata*, has been shortened for clarity.

Comparison of natural vs. synthetic allopolyploid plastomes and short-liguled populations of independent origin

Plastome sequences for the natural and synthetic long-liguled *T. miscellus* shared 100% identity (Table 3, Fig. 3). Similarly, the sequence of the synthetic short-liguled *T. miscellus* (111-7-5) shared 100% identity with all but one of the naturally formed short-liguled *T. miscellus* individuals. The exception was the short-liguled *T. miscellus* 2625-3, with which the synthetic short-liguled *T. miscellus* sequence shared 99.52% identity and differed by 206 polymorphisms (Table 3).

Phylogenetic analyses

The two phylogenetic analyses recovered topologies that were highly similar. Two main clades were recovered in both analyses (Fig. 3). One clade was well-supported (bootstrap support 100% in both analyses) and included the four representatives of *T. dubius*, the long-liguled *T. miscellus* individuals, and the short-liguled Albion *T. miscellus* (2625-3). The other main clade, which included both *T. pratensis* accessions and all but one of the short-liguled *T. miscellus* individuals, was well-supported (bootstrap support 96%) by the alignment that included a single copy of the IR, but more poorly so by the alignment comprising only coding sequences (bootstrap support <50%). The *T. dubius* clade consisted of two well-supported subclades (Fig. 3). One comprised the Chinese (OR840963) and Pullman *T. dubius* (2613-11, 2613-24) accessions as well as the long-liguled *T. miscellus* individuals (129-7-10, 2605-23, bootstrap support 97%–100%), and the other, the Spangle *T. dubius* (2616-1) and the Albion short-liguled *T. miscellus* (2625-3, bootstrap support 98%–100%). In the *T. pratensis* clade, the *T. pratensis* and short-liguled *T. miscellus* individuals from the United States formed a well-supported clade (bootstrap support 100% in both analyses) sister to the individual from the United Kingdom.

DISCUSSION

Divergence between diploid *Tragopogon* plastomes is low, but variation exists within and between species that may identify contributing progenitor plastomes in allopolyploids

The origins of the North American *Tragopogon* polyploids have been well-studied (e.g., Soltis *et al.* 2004; Symonds *et al.* 2010), but comparisons of the whole plastome sequences assembled in the current study have provided new insights. Specifically, despite the plastome sequences of the diploid individuals exhibiting high identity, sequence variation within and between *T. dubius* and *T. pratensis* may be sufficient to determine the specific origins of individual *T. miscellus* populations. For example, the naturally occurring long-liguled *T. miscellus* (2605-23) individual included in the present study was from Pullman, Washington, and it is therefore not surprising that the plastome sequence from this plant shared 100% identity with those of the two Pullman *T. dubius* individuals. Likewise, it is also not surprising that the synthetic long-liguled *T. miscellus* (129-7-10) plant shared 100% identity with the Pullman *T. dubius*, because the 2613-24 individual was the maternal progenitor for this synthetic lineage (Tate *et al.* 2009).

Although *T. pratensis* is more commonly the maternal progenitor of *T. miscellus* allopolyploids, it would be worthwhile surveying plastome diversity in *T. dubius* further as previous studies suggest that distinct nuclear (Soltis *et al.* 2022) and plastome (Soltis & Soltis 1989) genotypes are present in North America. Based on restriction enzyme recognition sequences and inferred fragment sizes, the two plastome genotypes identified in the present study seemingly correspond to those found by Soltis & Soltis (1989), who also sampled the Pullman and Spangle, Washington, populations. That said, the broader geographic sampling of the Soltis *et al.* (2022) study suggests that these genotypes, or variants of them, may be more widely distributed. Additionally, in our phylogenetic analysis, the *T. dubius* individual from China, where this species is considered invasive (Huang *et al.* 2024), was resolved as distinct from both the Pullman and Spangle *T. dubius* individuals (Fig. 3). There is clearly additional genetic variation within *T. dubius*, but the extent of plastome diversity in North America is uncertain. Given the well-documented genetic diversity and taxonomic complexity of this species (e.g., Mavrodiev *et al.* 2012), nuclear and plastid variation from across the range of *T. dubius* should be considered when assessing the origins of allopolyploid lineages. For example, long-liguled *T. miscellus* individuals have recently been found in Sheridan, Wyoming, and these appear to represent a novel origin (D. Soltis and P. Soltis, unpublished data).

We included just one North American *T. pratensis* individual here, sampling that likely underrepresents the plastome diversity

of this species in North America (Soltis *et al.* 2023). Indeed, individuals from North America and the United Kingdom were resolved as distinct in our phylogenetic analyses, indicating that plastome sequences vary within this species (Fig. 3). That said, given the wider geographic sampling of short-liguled *T. miscellus* in this study (Table 1), it is a little surprising that plastome sequences from all but one of the natural short-liguled *T. miscellus* individuals shared 100% identity with the *T. pratensis* from Moscow, Idaho (Table S1). That the synthetic short-liguled *T. miscellus* individual (111–7-5) shared the same plastome sequence as most of the sampled natural allopolyploids is less surprising given that its maternal progenitor was a *T. pratensis* from Moscow (2608-1) (Tate *et al.* 2009). Going forward it will be important to further assess plastome sequence diversity from across the ranges of *T. pratensis* and short-liguled *T. miscellus*. Investigating synthetic *T. miscellus* individuals with parents from different populations would also be worthwhile. Again, this would allow us to determine whether other plastome genotypes have been captured when geographically distinct *T. pratensis* parents were involved in the crossing. For example, *T. pratensis* from Spangle was the maternal parent for other synthetic lines (e.g., 67 and 79; Tate *et al.* 2009).

Short-liguled Albion *T. miscellus* 2625-3 is a back-crossed individual to *T. dubius*

As short-liguled allopolyploids, the two *T. miscellus* individuals sampled from Albion, Washington (2625-3 and 2625-6) were both expected to have plastome sequences consistent with those from representatives of *T. pratensis*. As expected, the plastome assembled for one of these individuals (2625-6) shared 100% identity with the North American *T. pratensis* plastome sequence. In contrast, the sequence from the other individual (2625-3) shared 99.98% identity with that of *T. dubius* (2616-1) from Spangle (Table S1, Fig. 3). We suggest that this latter *T. miscellus* individual (2625-3) arose as a result of a backcross with *T. dubius*.

Studies utilising various nuclear markers previously reported anomalies with *T. miscellus* individuals from Albion (Tate *et al.* 2009; Symonds *et al.* 2010; Sehrish *et al.* 2015). Using cleaved amplified polymorphic sequences (CAPS), Tate, Joshi, *et al.* (2009) showed an elevated number of homeolog losses (six) for *T. miscellus* (2625-3) relative to the other individuals from Albion (0–3) and other populations (usually one). Moreover, *T. miscellus* (2625-3) lost the *T. pratensis* homeologs, rather than the *T. dubius* homeologs as in other populations (Tate *et al.* 2009). Sehrish *et al.* (2015) evaluated cytonuclear compatibility using the Rubisco (*rbcS*/*rbcL*) system for the same samples. While most *T. miscellus* allopolyploids retained both parental nuclear *rbcS* homeologs, three individuals from Albion, including 2625-3 and 2625-6, lost the presumably maternal copy of *rbcS* that was expected to be retained (i.e., in order to maintain cytonuclear coordination) (Sehrish *et al.* 2015). In the case of *T. miscellus* (2625-3), backcrossing with *T. dubius* and inheritance of the *T. dubius* plastome might explain loss of the *T. pratensis* copy of *rbcS*. We assumed that this individual was an allotetraploid, but it is possible that it could have been a triploid.

Tragopogon miscellus (2625-3) is the only example in the current study of an allopolyploid with a plastome sequence that does not match exactly one of the plastome sequences from a diploid included here. Specifically, the *T. miscellus* (2625-3) plastome sequence differed from that of the closest diploid, *T. dubius* (2616-1), by five substitutions and six indels including five that involve homopolymer repeats. It is possible that these polymorphisms arose following allopolyploid formation, although given that none of the substitutions occurred within a gene, it seems unlikely that these have been driven by changes in cytonuclear interactions. Alternatively, because our sampling of *T. dubius* remains limited, these differences might be explained by additional, as yet unsampled, plastome diversity in the diploid. In this case, the backcrossing would have involved a *T. dubius* population with a distinct plastome sequence. This possibility is consistent with *T. dubius* exhibiting plastome diversity and the observation that the *T. dubius* (2616-1) and *T. miscellus* (2625-3) populations are geographically distinct (Table 1). Distinguishing between these options (i.e., novel mutations or unsampled parental diversity) requires further assessment of plastome diversity in *T. dubius*. Symonds *et al.* (2010) provides a potential starting point for such an assessment, suggesting shared ancestry between the Albion allopolyploids and Moscow populations.

Further investigations of the Albion *T. miscellus* population are hampered by its apparent extinction. The Albion *T. miscellus* population was likely one of the "small and precarious" populations described by Ownbey (1950) and was reported by Novak *et al.* (1991) to consist of 30–40 short-liguled *T. miscellus* plants, together with both parental and hybrid individuals. When the samples included in the present study were collected from Albion in 2004, field notes from this site included "*T. miscellus* and *T. dubius* together plus some weird flowers—hybrids?" (D. Soltis and P. Soltis, unpublished). Despite visits over several years since then, allopolyploids have not been seen again at this site, whereas *T. dubius* remains abundant in Albion and surrounding areas (D. Soltis, P. Soltis, and J. Tate, pers. obs.). It is not known what caused the loss of this allopolyploid population. Given the small size of the population, widespread disturbance at the site may be to blame or perhaps the polyploids were overcome genetically and/or ecologically by *T. dubius* as their numbers declined. In the latter case, the generation of triploids, which would be sterile, may have contributed to the decline. The other recently formed North American *Tragopogon* allopolyploid, *T. mirus*, also existed in Albion (Novak *et al.* 1991) but no longer occurs there. As surveys have previously shown, although allopolyploid *Tragopogon* populations have been lost, others have persisted and even expanded in size (Novak *et al.* 1991; Soltis *et al.* 2004), while still others have arisen *de novo* (Soltis *et al.* 2012).

Plastome variation: origins of allopolyploids and genome dynamics

Studies across angiosperms have shown that several plastome regions consistently exhibit elevated levels of sequence variation (e.g., Henriquez *et al.* 2020; Xu *et al.* 2024). Referred to as "hypervariable", these rapidly evolving regions are often targeted for barcoding or phylogenetic studies (Fazekas *et al.* 2008; Hollingsworth *et al.* 2011; Dong *et al.* 2012, 2015). Several of the regions previously recognised as mutational hot spots (Barthet & Hilu 2008; Ahmed *et al.* 2012; Dong *et al.* 2015; Amar 2020) were also polymorphic in *Tragopogon*, including the *matK*, *ndhF*, *ycf1*, and *ycf2* genes as well as the *rpl16* intron. In the current analyses, these regions vary both

within and between the diploid progenitor species. Although more work is needed to fully characterise sequence diversity at these loci, our results suggest that targeted sequencing of specific loci may be sufficient to identify the putative maternal parent to at least the species level. Further, given that plastomes evolve relatively slowly and that these *Tragopogon* allopolyploids arose within the last 100 years, it is perhaps not surprising that the allopolyploids themselves could not be distinguished from their maternal parents at these rapidly evolving loci or elsewhere along the plastome.

Given the more dynamic nature of the nuclear genome, it may be that changes to cytonuclear interactions in recently formed allopolyploids can be accommodated more readily by changes to the nuclear component. Yet, investigating the consequences of substitutions in plastome gene sequences may help to identify novel cytonuclear interactions. In the present study, 24 nonsynonymous substitutions were identified between the two diploid progenitor species (Table 4). These are potential candidates for investigating the maintenance of cytonuclear coordination in the allopolyploids. Of particular interest are substitutions that change the polarity of the encoded amino acid as these may alter how the resulting proteins interact (Moreira *et al.* 2007). Three such changes were identified in comparisons of *T. dubius* from Pullman (2613-24) and *T. pratensis* (2608-35) from Moscow, but additional differences may exist in other diploid populations (Soltis *et al.* 2022, 2023).

Synonymous SNPs may have unknown phenotypic effects in the absence of other evidence (Vihinen 2022). For example, Sehrish *et al.* (2015) identified only a single synonymous SNP between plastid *rbcL* regions of *T. dubius* and *T. pratensis*, but most *T. miscellus* individuals showed differential expression patterns of the interacting nuclear *rbcS* homeologs that favoured the maternally inherited copy. Thus, some of these synonymous SNPs may be important in facilitating cytonuclear coordination.

CONCLUSIONS

Although studies of allopolyploid evolution often focus on the duplicated nuclear genomes, understanding the sources and evolutionary dynamics of the plastid and mitochondrial genomes may provide novel insights. The *Tragopogon* allopolyploids investigated here are an interesting case study, as despite their recent origins, there is sufficient plastome diversity to address questions about their parental contributions and evolution. In other plant groups, polyploids may be hundreds to perhaps millions of years older than in *Tragopogon*, and in such cases, we would expect plastome diversity to be of even greater utility.

AUTHOR CONTRIBUTIONS

UM: Writing—original draft, methodology, formal analysis, data curation. SCN: Methodology, formal analysis, writing—review and editing. RCW: Methodology, formal analysis, supervision, writing—review and editing. PSS: Conceptualization, writing—review and editing. DES: Conceptualization, writing—review and editing. JAT: Conceptualization, funding acquisition, methodology, data curation, project administration, supervision, writing—review and editing.

ACKNOWLEDGEMENTS

Funding for this project came from a Massey University Research Fund to JT and a Royal Society of New Zealand Marsden Fund (18-MAU-049) to JT, DS, and PS. This work was also supported by grant DEB-2043478 from the U.S. National Science Foundation to DS and PS. We are grateful to Xiaoxiao Lin at the Massey Genome Service for advice on the project and facilitating sequencing of the plastomes. Open access publishing was facilitated by Massey University, as part of the Wiley - Massey University agreement via the Council of Australian University Librarians. Open access publishing was facilitated by Massey University, as part of the Wiley - Massey University agreement via the Council of Australian University Librarians.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. Percent identities (above diagonal) and total nucleotide differences (below diagonal) for the 12 *Tragopogon* plastomes.

REFERENCES

Ahmed I., Biggs P.J., Matthews P.J., Collins L.J., Hendy M.D., Lockhart P.J. (2012) Mutational dynamics of aroid chloroplast genomes. *Genome Biology and Evolution*, **4**, 1316–1323.

Altschul S.F., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, **25**, 3389–3402.

Amar M.H. (2020) *ycf1-ndhF* genes, the most promising plastid genomic barcode, sheds light on phylogeny at low taxonomic levels in *Prunus persica*. *Journal of Genetic Engineering and Biotechnology*, **18**, 42.

Ashton P.A., Abbott R.J. (1992) Multiple origins and genetic diversity in the newly arisen allopolyploid species, *Senecio cambrensis* Rosser (Compositae). *Heredity*, **68**, 25–32.

Barker M.S., Arrigo N., Baniaga A.E., Li Z., Levin D.A. (2016) On the relative abundance of autopolyploids and allopolyploids. *New Phytologist*, **210**, 391–398.

Barthet M.M., Hilu K.W. (2008) Evaluating evolutionary constraint on the rapidly evolving gene *matK* using protein composition. *Journal of Molecular Evolution*, **66**, 85–97.

Brochmann C., Soltis P.S., Soltis D.E. (1992a) Multiple origins of the octoploid Scandinavian endemic *Draba cacuminum*: Electrophoretic and morphological evidence. *Nordic Journal of Botany*, **12**, 257–272.

Brochmann C., Soltis P.S., Soltis D.E. (1992b) Recurrent formation and polyphyly of Nordic polyploids in *Draba* (Brassicaceae). *American Journal of Botany*, **79**, 673–688.

Buggs R.J.A., Chamala S., Wu W., Tate J.A., Schnable P.S., Soltis D.E., Soltis P.S., Barbazuk W.B. (2012) Rapid, repeated, and clustered loss of duplicate genes in allopolyploid plant populations of independent origin. *Current Biology*, **22**, 248–252.

Buggs R.J.A., Doust A.N., Tate J.A., Koh J., Soltis K., Feltus F.A., Paterson A.H., Soltis P.S., Soltis D.E. (2009) Gene loss and silencing in *Tragopogon miscellus* (Asteraceae): Comparison of natural and synthetic allotetraploids. *Heredity*, **103**, 73–81.

Buggs R.J.A., Zhang L., Miles N., Tate J.A., Gao L., Schnable P.S., Buggs R.J., Wei W., Barbazuk W.B., Soltis P.S., Soltis D.E. (2011) Transcriptomic shock generates evolutionary novelty in a newly formed, natural allopolyploid plant. *Current Biology*, **21**, 551–556.

Chen N., Chen W.J., Yan H., Wang Y., Kang H.Y., Zhang H.Q., Zhou Y.H., Sun G.L., Sha L.N., Fan X. (2020) Evolutionary patterns of plastome uncover diploid-polyploid maternal relationships in Triticeae. *Molecular Phylogenetics and Evolution*, **149**, 106838.

Chen S., Yan H., Sha L., Chen N., Zhang H., Zhou Y., Fan X. (2021) Chloroplast phylogenomic analyses resolve multiple origins of the *Kengyilia* species

(Poaceae: Triticeae) via independent polyploidization events. *Frontiers in Plant Science*, **12**, 682040.
Chen Z.J., Ni Z.F. (2006) Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. *BioEssays*, **28**, 240–252.
Cheng F., Sun C., Wu J., Schnable J., Woodhouse M.R., Liang J., Cai C., Freeling M., Wang X. (2016) Epigenetic regulation of subgenome dominance following whole genome triplication in *Brassica rapa*. *New Phytologist*, **211**, 288–299.
Chernomor O., Von Haeseler A., Minh B.Q. (2016) Terrace aware data structure for phylogenomic inference from supermatrices. *Systematic Biology*, **65**, 997–1008.
Cook L.M., Soltis P.S., Brunsfeld S.J., Soltis D.E. (1998) Multiple independent formations of *Tragopogon* tetraploids (Asteraceae): Evidence from RAPD markers. *Molecular Ecology*, **7**, 1293–1302.
Corriveau J.L., Coleman A.W. (1988) Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *American Journal of Botany*, **75**, 1443–1458.
Dong W., Liu J., Yu J., Wang L., Zhou S. (2012) Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PLoS One*, **7**, e35071.
Dong W., Xu C., Li C., Sun J., Zuo Y., Shi S., Cheng T., Guo J., Zhou S. (2015) *ycf1*, the most promising plastid DNA barcode of land plants. *Scientific Reports*, **5**, 8348.
Doyle J.J., Coate J.E. (2019) Polyploidy, the nucleotype, and novelty: The impact of genome doubling on the biology of the cell. *International Journal of Plant Sciences*, **180**, 1–52.
Doyle J.J., Doyle J.L. (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, **19**, 11–15.
Doyle J.J., Doyle J.L., Brown A.H.D., Palmer R.G. (2002) Genomes, multiple origins, and lineage recombination in the *Glycine tomentella* (Leguminosae) polyploid complex: Histone H3-D gene sequences. *Evolution*, **56**, 1388–1402.
Fazekas A.J., Burgess K.S., Kesanakurti P.R., Graham S.W., Newmaster S.G., Husband B.C., Percy D.M., Hajibabaei M., Barrett S.C.H. (2008) Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS One*, **3**, e2802.
Gornicki P., Zhu H., Wang J., Challa G.S., Zhang Z., Gill B.S., Li W. (2014) The chloroplast view of the evolution of polyploid wheat. *New Phytologist*, **204**, 704–714.
Henriquez C.L., Abdullah, Ahmed I., Carlsen M.M., Zuluaga A., Croat T.B., McKain M.R. (2020) Evolutionary dynamics of chloroplast genomes in subfamily Aroideae (Araceae). *Genomics*, **112**, 2349–2360.
Heslop-Harrison J.S., Schwarzacher T., Liu Q. (2022) Polyploidy: Its consequences and enabling role in plant diversification and evolution. *Annals of Botany*, **131**, 1–10.
Hollingsworth P.M., Graham S.W., Little D.P. (2011) Choosing and using a plant DNA barcode. *PLoS One*, **6**, e19254.
Huang Y., Cao J., Zhao M., Guo J., Li J., Wang R. (2024) The complete chloroplast genome sequence of an invasive plant, *Tragopogon dubius* Scopoli (Asteraceae). *Mitochondrial DNA Part B Resources*, **9**, 352–356.
Jiao Y., Wickett N.J., Ayyampalayam S., Chanderbali A.S., Landherr L., Ralph P.E., Tomsho L.P., Hu Y., Liang H., Soltis P.S., Soltis D.E., Clifton S.W., Schlarbaum S.E., Schuster S.C., Ma H., Leebens-Mack J., dePamphilis C.W. (2011) Ancestral polyploidy in seed plants and angiosperms. *Nature*, **473**, 97–100.
Jin J.-J., Yu W.-B., Yang J.-B., Song Y., dePamphilis C.W., Yi T.-S., Li D.Z. (2020) GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology*, **21**, 241.
Kalyaanamoorthy S., Wong T.K.F., Jermiin L.S., Minh B.Q., Von Haeseler A. (2017) ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, **14**, 587–589.
Kan J., Nie L., Wang M., Tiwari R., Tembrock L.R., Wang J. (2024) The mendelian pea pan-plastome: Insights into genomic structure, evolutionary history, and genetic diversity of an essential food crop. *Genomics Communications*, **1**, e004.
Katoh K., Misawa K., Kuma K.i., Miyata T. (2002) MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, **30**, 3059–3066.
Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A., Markowitz S., Duran C., Thierer T., Ashton B., Meintjes P., Drummond A. (2012) Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647–1649.
Koh J., Soltis P.S., Soltis D.E. (2010) Homeolog loss and expression changes in natural populations of the recently and repeatedly formed allotetraploid *Tragopogon mirus* (Asteraceae). *BMC Genomics*, **11**, 97.
Levin D.A. (2003) The cytoplasmic factor in plant speciation. *Systematic Botany*, **28**, 5–11.
Li H., Durbin R. (2009) Making the leap: Maq to BWA. *Mass Genomics*, **25**, 1754–1760.
Lohse M., Drechsel O., Kahlau S., Bock R. (2013) OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Research*, **41**, W575–W581.
Magdy M., Ou L., Yu H., Chen R., Zhou Y., Hassan H., Feng B., Taitano N., van der Knaap E., Zou X., Li F., Ouyang B. (2019) Pan-plastome approach empowers the assessment of genetic variation in cultivated *Capsicum* species. *Horticulture Research*, **6**, 108.
Mavrodiev E.V., Gitzendanner M., Calaminus A.K., Baldini R.M., Soltis P.S., Soltis D.E. (2012) Molecular phylogeny of *Tragopogon* L. (Asteraceae) based on seven nuclear loci (*Adh*, *GapC*, *LFY*, *AP3*, *PI*, ITS, and ETS). *Webbia*, **67**, 111–137.
Milne I., Bayer M., Cardle L., Shaw P., Stephen G., Wright F., Marshall D. (2009) Tablet—Next generation sequence assembly visualization. *Bioinformatics*, **26**, 401–402.
Minh B.Q., Lanfear R., Schmidt H.A., Chernomor O., Schrempf D., Von Haeseler A., Woodhams M.D. (2020) IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, **37**, 1530–1534.
Moreira I.S., Fernandes P.A., Ramos M.J. (2007) Hot spots—A review of the protein–protein interface determinant amino-acid residues. *Proteins: Structure, Function, and Bioinformatics*, **68**, 803–812.
Mower J.P., Sloan D.B., Alverson A. (2012) Plant mitochondrial diversity: The genomics revolution. In: Wendel J., Greilhuber J., Dolezel J., Leitch I.J. (Eds), *Plant genome diversity volume 1: Plant genomes, their residents, and their evolutionary dynamics*. Springer Verlag, Wien, pp 123–144.
Novak S.J., Soltis D.E., Soltis P.S. (1991) Ownbey's Tragopogons: 40 years later. *American Journal of Botany*, **78**, 1586–1600.
Ownbey M. (1950) Natural hybridization and amphiploidy in the genus *Tragopogon*. *American Journal of Botany*, **37**, 487–499.
Ownbey M., McCollum G.D. (1953) Cytoplasmic inheritance and reciprocal amphiploidy in *Tragopogon*. *American Journal of Botany*, **40**, 788–796.
Peng Y., Leung H.C.M., Yiu S.M., Chin F.Y.L. (2012) IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*, **28**, 1420–1428.
Pennisi E. (2016) Do genomic conflicts drive evolution? *Science*, **353**, 334–335.
Ramsey J., Schemske D.W. (1998) Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics*, **29**, 467–501.
Rand D.M., Haney R.A., Fry A.J. (2004) Cytonuclear coevolution: The genomics of cooperation. *Trends in Ecology & Evolution*, **19**, 645–653.
Ruhlman T.A., Jansen R.K. (2014) The plastid genomes of flowering plants. In: Maliga P. (Ed), *Chloroplast biotechnology: Methods and protocols*. Humana Press, Totowa, NJ, pp 3–38.
Segraves K.A., Thompson J.N., Soltis P.S., Soltis D.E. (1999) Multiple origins of polyploidy and the geographic structure of *Heuchera grossulariifolia*. *Molecular Ecology*, **8**, 253–262.
Sehrish T., Symonds V.V., Soltis D.E., Soltis P.S., Tate J.A. (2014) Gene silencing via DNA methylation in naturally occurring *Tragopogon miscellus* (Asteraceae) allopolyploids. *BMC Genomics*, **15**, 701.
Sehrish T., Symonds V.V., Soltis D.E., Soltis P.S., Tate J.A. (2015) Cytonuclear coordination is not immediate upon allopolyploid formation in *Tragopogon miscellus* (Asteraceae) allopolyploids. *PLoS One*, **10**, e0144339.
Shan S., Boatwright J.L., Liu X., Chanderbali A.S., Fu C., Soltis P.S., Soltis D.E. (2020) Transcriptome dynamics of the inflorescence in reciprocally formed allopolyploid *Tragopogon miscellus* (Asteraceae). *Frontiers in Genetics*, **11**, 888.
Shan S., Spoelhof J.P., Blischak P.D., Batley J., Soltis P.S., Soltis D.E., Edwards D. (2025) Pangenomes provide new insights into polyploidy in plants. *Evolutionary Journal of the Linnean Society*, **4**, kzaf010.
Sharbrough J., Conover J.L., Tate J.A., Wendel J.F., Sloan D.B. (2017) Cytonuclear responses to genome doubling. *American Journal of Botany*, **104**, 1277–1280.
Sloan D.B. (2015) Using plants to elucidate the mechanisms of cytonuclear co-evolution. *New Phytologist*, **205**, 1040–1046.
Sloan D.B., Conover J.L., Grover C.E., Wendel J.F., Sharbrough J. (2024) Polyploid plants take cytonuclear perturbations in stride. *The Plant Cell*, **36**, 829–839.
Soltis D.E., Buggs R.J.A., Barbazuk W.B., Chamala S., Chester M., Gallagher J.P., Schnable P.S., Soltis P.S. (2012) Rapid and repeated evolution in the early stages of polyploidy: Genomic and cytogenetic studies of recent polyploidy in *Tragopogon*. In: Soltis P.S., Soltis D.E. (Eds), *Polyploidy and genome evolution*. Springer Berlin, Heidelberg, pp 271–292.
Soltis D.E., Mavrodiev E.V., Brukhin V., Roalson E.H., Albach D.C., Godden G.T., Alexeev Y.E.,

Gitzendanner M.A., Freeman C.C., Suárez-Santiago V.N., Soltis P.S. (2023) *Tragopogon pratensis*: Multiple introductions to North America, circumscription, and the formation of the allotetraploid *T. miscellus*. *Taxon*, **72**, 848–861.

Soltis D.E., Mavrodiev E.V., Gitzendanner M.A., Alexeev Y.E., Godden G.T., Soltis P.S. (2022) *Tragopogon dubius*: Multiple introductions to North America and the formation of the New World tetraploids. *Taxon*, **71**, 1287–1298.

Soltis D.E., Mavrodiev E.V., Myers S.C., Severns P.M., Zhang L., Gitzendanner M.A., Ayers T., Chester M., Soltis P.S. (2012) Additional origins of Ownbey's *Tragopogon mirus*. *Botanical Journal of the Linnean Society*, **169**, 297–311.

Soltis D.E., Soltis P.S. (1989) Allopolyploid speciation in *Tragopogon*: Insights from chloroplast DNA. *American Journal of Botany*, **76**, 1119–1124.

Soltis D.E., Soltis P.S. (1999) Polyploidy: Recurrent formation and genome evolution. *Trends in Ecology & Evolution*, **14**, 348–352.

Soltis D.E., Soltis P.S., Pires J.C., Kovarik A., Tate J.A., Mavrodiev E. (2004) Recent and recurrent polyploidy in *Tragopogon* (Asteraceae): Cytogenetic, genomic and genetic comparisons. *Biological Journal of the Linnean Society*, **82**, 485–501.

Soltis D.E., Visger C.J., Marchant D.B., Soltis P.S. (2016) Polypoidy: Pitfalls and paths to a paradigm. *American Journal of Botany*, **103**, 1–21.

Soltis P.S., Soltis D.E. (2000) The role of genetic and genomic attributes in the success of polyploids. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 7051–7057.

Stebbins G.L. (1947) Types of polyploids: Their classification and significance. *Advances in Genetics*, **1**, 403–429.

Straub S.C.K., Parks M., Weitemier K., Fishbein M., Cronn R.C., Liston A. (2012) Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *American Journal of Botany*, **99**, 349–364.

Symonds V.V., Soltis P.S., Soltis D.E. (2010) Dynamics of polyploid formation in *Tragopogon* (Asteraceae): Recurrent formation, gene flow, and population structure. *Evolution*, **64**, 1984–2003.

Tate J.A., Joshi P., Soltis K.A., Soltis P.S., Soltis D.E. (2009) On the road to diploidization? Homoeolog loss in independently formed populations of the allopolyploid *Tragopogon miscellus* (Asteraceae). *BMC Plant Biology*, **9**, 80.

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

Tate J.A., Symonds V.V., Doust A.N., Buggs R.J.A., Mavrodiev E.V., Majure L.C., Mavrodiev E., Soltis P.S., Soltis D.E. (2009) Synthetic polyploids of *Tragopogon miscellus* and *T. mirus* (Asteraceae): 60 years after Ownbey's discovery. *American Journal of Botany*, **96**, 979–988.

Tillich M., Lehwark P., Pellizzer T., Ulbricht-Jones E.S., Fischer A., Bock R., Greiner S. (2017) GeSeq – Versatile and accurate annotation of organelle genomes. *Nucleic Acids Research*, **45**, W6–W11.

Twyford A.D., Ness R.W. (2017) Strategies for complete plastid genome sequencing. *Molecular Ecology Resources*, **17**, 858–868.

Vamosi J.C., Dickinson T.A. (2006) Polyploidy and diversification: A phylogenetic investigation in Rosaceae. *International Journal of Plant Sciences*, **167**, 349–358.

Van de Peer Y., Ashman T.-L., Soltis P.S., Soltis D.E. (2020) Polyploidy: An evolutionary and ecological force in stressful times. *The Plant Cell*, **33**, 11–26.

Vihinen M. (2022) When a synonymous variant is nonsynonymous. *Genes*, **13**, 1485.

Wang J., Kan J., Wang J., Yan X., Li Y., Soe T., Tembrock L.R., Xing G., Li S., Wu Z., Jia M. (2024) The pan-plastome of *Prunus mume*: Insights into *Prunus* diversity, phylogeny, and domestication history. *Frontiers in Plant Science*, **15**, 1404071.

Wang S., Gao J., Chao H., Li Z., Pu W., Wang Y., Chen M. (2022) Comparative chloroplast genomes of *Nicotiana* species (Solanaceae): Insights into the genetic variation, phylogenetic relationship, and polyploid speciation. *Frontiers in Plant Science*, **13**, 899252.

Wendel J.F. (2015) The wondrous cycles of polyploidy in plants. *American Journal of Botany*, **102**, 1753–1756.

Wicke S., Schneeweiss G.M., dePamphilis C.W., Muller K.F., Quandt D. (2011) The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Molecular Biology*, **76**, 273–297.

Xu Y., Li Y., Chen Y., Wang L., Xue B., Zhang X., Song W., Guo W., Wu W. (2024) Comparative analysis of complete chloroplast genomes of *Rubus* in China: Hypervariable regions and phylogenetic relationships. *Genes*, **15**, 716.

Yan X.-L., Kan S.-L., Wang M.-X., Li Y.-Y., Tembrock L.R., He W.-C., Yan X.-L., Kan S.-L., Wang M.-X., Li Y.-Y., He W.-C., Nie L.-Y., Hu G.-J., Yuan D.-J., Ma X.-F., Wu Z.-Q. (2024) Genetic diversity and evolution of the plastome in allotetraploid cotton (*Gossypium* spp.). *Journal of Systematics and Evolution*, **62**, 1118–1136.

Yoo M.J., Koh J., Boatwright J.L., Soltis D.E., Soltis P.S., Barbazuk W.B., Chen S. (2024) Investigation of regulatory divergence between homoeologs in the recently formed allopolyploids, *Tragopogon mirus* and *T. miscellus* (Asteraceae). *The Plant Journal*, **117**, 1191–1205.

Yoo M.J., Szadkowski E., Wendel J.F. (2013) Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity*, **110**, 171–180.

Zuo S., Guo X., Mandáková T., Edginton M., Al-Shehbaz I.A., Lysak M.A. (2022) Genome diploidization associates with cladogenesis, trait disparity, and plastid gene evolution. *Plant Physiology*, **190**, 403–420.