New Phytologist

Research

Genome-wide DNA methylation dynamics following recent polyploidy in the allotetraploid *Tragopogon miscellus* (Asteraceae)

Shengchen Shan[1], Matthew A. Gitzendanner[2], J. Lucas Boatwright[3], Jonathan P. Spoelhof[1], Christina L. Ethridge[4], Lexiang Ji[5], Xiaoxian Liu[1,2,6], Pamela S. Soltis[1,7,8], Robert J. Schmitz[4] and Douglas E. Soltis[1,2,7,8]

[1]Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA; [2]Department of Biology, University of Florida, Gainesville, FL 32611, USA; [3]Advanced Plant Technology Program, Clemson University, Clemson, SC 29634, USA; [4]Department of Genetics, University of Georgia, Athens, GA 30602, USA; [5]Institute of Bioinformatics, University of Georgia, Athens, GA 30602, USA; [6]Bioinformatics Core, H. Lee Moffitt Cancer Center & Research Institute, Tampa, FL 33612, USA; [7]Biodiversity Institute, University of Florida, Gainesville, FL 32611, USA; [8]Genetics Institute, University of Florida, Gainesville, FL 32610, USA

Author for correspondence:
Shengchen Shan
Email: shan158538@ufl.edu

Received: *18 September 2023*
Accepted: *15 January 2024*

New Phytologist (2024)
doi: 10.1111/nph.19655

Key words: differentially methylated genes, differentially methylated regions, nonadditive methylation, polyploidy, *Tragopogon*, transposable elements.

Summary

- Polyploidy is an important evolutionary force, yet epigenetic mechanisms, such as DNA methylation, that regulate genome-wide expression of duplicated genes remain largely unknown. Here, we use *Tragopogon* (Asteraceae) as a model system to discover patterns and temporal dynamics of DNA methylation in recently formed polyploids.
- The naturally occurring allotetraploid *Tragopogon miscellus* formed in the last 95–100 yr from parental diploids *Tragopogon dubius* and *T. pratensis*. We profiled the DNA methylomes of these three species using whole-genome bisulfite sequencing.
- Genome-wide methylation levels in *T. miscellus* were intermediate between its diploid parents. However, nonadditive CG and CHG methylation occurred in transposable elements (TEs), with variation among TE types. Most differentially methylated regions (DMRs) showed parental legacy, but some novel DMRs were detected in the polyploid. Differentially methylated genes (DMGs) were also identified and characterized.
- This study provides the first assessment of both overall and locus-specific patterns of DNA methylation in a recent natural allopolyploid and shows that novel methylation variants can be generated rapidly after polyploid formation. Together, these results demonstrate that mechanisms to regulate duplicate gene expression may arise soon after allopolyploid formation and that these mechanisms vary among genes.

Introduction

Polyploidy (i.e. whole-genome duplication, WGD) is a major evolutionary mechanism in plants (e.g. Soltis *et al.*, 2015; Van de Peer *et al.*, 2021), and it contributes to genomic, transcriptomic, and epigenomic diversity while being associated with phenotypic innovations and increased evolvability and diversification (Chen, 2007; Soltis *et al.*, 2014; Soltis & Soltis, 2016; Edger *et al.*, 2017; Landis *et al.*, 2018; Doyle & Coate, 2019; Fox *et al.*, 2020; Van de Peer *et al.*, 2021). Wood *et al.* (2009) estimated that 35% of extant vascular plants may have originated via polyploidy, and all living angiosperms experienced at least one WGD during their evolution (Jiao *et al.*, 2011). In addition, many crops are relatively recent polyploids, having formed during the past 10 000 yr: for example, *Brassica napus* (4*x*) and bread wheat (6*x*) are *c.* 7500 and *c.* 8000 yr old, respectively (Haudry *et al.*, 2007; Chalhoub *et al.*, 2014). A better grasp of polyploid genome function and evolution is therefore needed for both understanding plant diversity and improving crop breeding (Renny-Byfield & Wendel, 2014). Central to polyploid genome function is the regulation and expression of duplicate genes. Furthermore, because phenotypic divergence of polyploids from their diploid parents may be due to differential gene expression, the factors that control gene expression – such as DNA methylation and other epigenetic mechanisms – may actually control phenotypic diversity.

DNA methylation, the addition of a methyl group to cytosine, plays a significant role in heterochromatin formation, silencing of transposable elements (TEs) and other repeated sequences, and sometimes regulation of gene expression (Law & Jacobsen, 2010). In green plants, DNA methylation occurs in three cytosine contexts: CG, CHG (H represents A, T, or C), and CHH. DNA methylation state is determined by the dynamics of *de novo* methylation, methylation maintenance, and demethylation. The molecular underpinnings of DNA methylation dynamics are well-understood in the model organism *Arabidopsis thaliana* (e.g. Cao & Jacobsen, 2002; Gong *et al.*, 2002; Cuerda-Gil &

Slotkin, 2016). However, such data are largely absent for other plant species (Niederhuth & Schmitz, 2017; Kumar & Mohapatra, 2021; Nie, 2021).

Flowering plant genomes are partitioned into unique DNA methylation states. Sequences that are transcriptionally silenced by DNA methylation are highly enriched for TEs and other repeated sequences and are typically methylated in the CG, CHG, and CHH sequence contexts (Cokus *et al.*, 2008; Lister *et al.*, 2008). Genes are present in three distinct classes: unmethylated, gene body DNA methylation, and transposon-like methylation (Kawakatsu *et al.*, 2016; Zhang *et al.*, 2020). In fact, most exons in flowering plant genomes are unmethylated (Niederhuth *et al.*, 2016). Gene body DNA methylation is unique in that it is present in moderately expressed 'housekeeping' genes with a yet-undefined function (Tran *et al.*, 2005; Zhang *et al.*, 2006; Zilberman, 2017). Transposon-like methylation genes show hallmarks of silenced transposons and repeats and are transcriptionally inactive (Kawakatsu *et al.*, 2016).

Hybridization and allopolyploidization both result in altered DNA methylation patterns (e.g. Lukens *et al.*, 2006; Parisod *et al.*, 2009; Hegarty *et al.*, 2011; Greaves *et al.*, 2012; Zhang *et al.*, 2016; Edger *et al.*, 2017; Li *et al.*, 2019). In allopolyploids, different TE methylation patterns between subgenomes possibly result in differentially expressed neighboring homeologs, and, in the long term, this could account for biased gene loss between parental genomes (Wendel *et al.*, 2018). Additionally, DNA methylation changes are also observed in autopolyploids (e.g. Zhang *et al.*, 2015; Hao *et al.*, 2023).

Most previous studies have used synthetic polyploids to investigate the impact of WGD on DNA methylation in the early generations following polyploidy. What are the immediate consequences of natural allopolyploidy on DNA methylation? There are few recently formed (< 200 yr old) natural polyploids with clear parentage (Soltis & Soltis, 2009; Wendel *et al.*, 2018, e.g. in *Spartina*, *Senecio*, *Cardamine*, *Tragopogon*, and *Mimulus*) in which to study the impact of polyploidy on DNA methylation shortly after polyploid formation. To date, patterns of DNA methylation have been investigated only in *Spartina* (Salmon *et al.*, 2005; Parisod *et al.*, 2009) and *Mimulus* (Edger *et al.*, 2017), and each analysis has limitations. For example, research in *Spartina* used the methylation-sensitive amplified polymorphism (MSAP) technique, which detects DNA methylation status at randomly distributed and anonymous loci but cannot detect cytosine methylation in the CHH context (Agius *et al.*, 2023). Using high-throughput whole-genome bisulfite sequencing (WGBS), Edger *et al.* (2017) compared the overall levels of genome-wide methylation at gene and TE regions (i.e. the metaplot analysis) in the < 140-yr-old naturally formed allopolyploid *Mimulus peregrinus*, its progenitor species, hybrids between the progenitors, and the resynthesized allopolyploid. Despite providing the most comprehensive insight into methylation in a young polyploid system to date, locus-specific patterns of DNA methylation changes following polyploidy were not investigated in *Mimulus*, and many fundamental questions remain: What are the differentially methylated regions (DMRs) between the parents? What are the DMRs between different subgenomes in the polyploid? Which DMRs exhibit parental legacy, and which are induced by polyploidy? What are the functional annotations of these DMRs? Furthermore, the impact of polyploidy on DNA methylation across TE types was not examined in *Mimulus*. Does the pattern observed in one type of TE apply to others? Despite the importance of WGD, a high-throughput and comprehensive analysis of methylome dynamics in the early stages following polyploid formation is lacking for any natural system, and generalities remain elusive. Moreover, because methylation patterns differ widely across diploid species (e.g. Niederhuth *et al.*, 2016), the effect of polyploidy – particularly allopolyploidy – on methylation will likely also vary to some extent among species. Thus, analyses of methylation in additional well-characterized polyploid complexes are necessary to develop a general framework for assessing the impact of methylation on genes duplicated by polyploidy and the timeline in which alterations develop.

One of the best-studied recently formed natural polyploid systems involves two polyploids in the genus *Tragopogon* (Asteraceae). The allotetraploids *Tragopogon miscellus* and *Tragopogon mirus* formed within the last *c.* 95–100 yr with clearly documented parentage (Ownbey, 1950; Tate *et al.*, 2009; Soltis *et al.*, 2012). The diploid parents of *T. miscellus* are *Tragopogon dubius* and *Tragopogon pratensis*, and those of *T. mirus* are *T. dubius* and *T. porrifolius*. Both *T. miscellus* and *T. mirus* formed repeatedly – at least 21 and 11 times, respectively (Soltis *et al.*, 1995, 2004; Symonds *et al.*, 2010). In addition, *T. miscellus* formed reciprocally with resultant distinct floral and inflorescence morphology: Those allotetraploids with *T. dubius* as the maternal parent have long ligules, and those with *T. pratensis* as the maternal parent have short ligules (Ownbey, 1950; Soltis & Soltis, 1989). This *Tragopogon* system, with two recently and recurrently formed allotetraploid species, provides an excellent model for studying the consequences of allopolyploidy across genetic and genomic scales. Previous studies have demonstrated that allopolyploidization has generated novel arrays of karyotypes, gene content, and gene expression among individual plants of both *T. mirus* and *T. miscellus*. For example, both naturally occurring and synthetic *T. miscellus* and *T. mirus* exhibit true aneuploidy (e.g. $2n = 23$ or 25 rather than the expected and typical $2n = 24$), compensated aneuploidy (in which $2n = 24$ but with unequal doses of some parental chromosomes), and intergenomic translocations (Lim *et al.*, 2008; Chester *et al.*, 2012, 2015; Spoelhof *et al.*, 2017). In addition, dynamic genetic changes, including gene loss and gene silencing, in the two newly formed *Tragopogon* polyploids have been well documented (e.g. Tate *et al.*, 2006, 2009; Buggs *et al.*, 2009, 2011, 2012; Boatwright *et al.*, 2018; Shan *et al.*, 2020), in all cases demonstrating both parental and novel patterns and variation among individuals. However, the epigenetic consequences of polyploidy in *Tragopogon* remain unclear. In the single study conducted to date, Sehrish *et al.* (2014) examined the DNA methylation status of four loci in *T. miscellus* and revealed that some silenced homeologs are methylated in the polyploids. Here, we examined genome-wide DNA methylation patterns in the allopolyploid *T. miscellus* (the short-liguled form) and its diploid parents, *T. dubius* and *T. pratensis*. We specifically addressed the aspects of methylation that

have not been addressed in other young polyploid systems: Is DNA methylation in *T. miscellus* additive of parental patterns? Does DNA methylation vary across loci, gene regions, types of TEs, and cytosine contexts? How do patterns of parental DNA methylation compare with methylation of the subgenomes in *T. miscellus*? What is the relative extent of parental legacy vs novelty in DNA methylation? Addressing these questions will clarify whether or not allopolyploidy can quickly lead to new epigenetic variants, with potential downstream impacts on gene expression and phenotypic diversity.

Materials and Methods

Plant material and DNA extraction

Tragopogon L. plants were grown in the Department of Biology glasshouse, University of Florida (Gainesville, FL, USA). Leaf material from 2-wk-old seedlings was collected and immediately frozen in liquid nitrogen. This study includes two diploid species (*T. dubius* Scop. and *T. pratensis* L.) and their allotetraploid derivative (*T. miscellus* Ownbey) with two replicate plants per species. As described previously, *T. miscellus* has originated multiple times, typically whenever the two diploid parental species co-occur, as in Garfield, WA, USA, which likely represents a site of polyploid formation (Soltis *et al.*, 1995, 2004; Symonds *et al.*, 2010). The plants used here were carefully chosen, to the extent possible, to reflect the genotypes involved in the origin of this allopolyploid population so that comparisons between *T. miscellus* and its diploid parental species did not incorporate genetic variants from other diploid-polyploid lineages (Symonds *et al.*, 2010; Soltis *et al.*, 2022, 2023). Leaf material of *T. miscellus* was collected from individuals 3059-7-7 and 3059-21-5, both from Garfield, WA, USA. For *T. pratensis*, both replicates were from Garfield, WA, USA (3058-1-2 and 3058-4-10). For diploid *T. dubius*, one replicate was from Garfield, WA, USA (3060-1-4), and the other individual was from Pullman, WA, USA (3040-6-2); even though this individual was not sampled in Garfield, it has the same ITS genotype as the Garfield *T. dubius* and likely represents the same lineage (Soltis *et al.*, 2022). DNA was extracted using the DNeasy Plant Mini Kit from Qiagen (Hilden, Germany). Herbarium vouchers for all individuals were deposited in the Florida Museum of Natural History Herbarium (FLAS).

MethylC-seq library preparation and high-throughput sequencing

The method of MethylC-seq library preparation followed that described in Urich *et al.* (2015). Briefly, the DNA samples (two replicates of each species, as described previously) were sonicated into *c.* 200-bp fragments. The EZ DNA Methylation-Gold Kit from Zymo (Irvine, CA, USA) was used for bisulfite conversion of genomic DNA to differentiate unmethylated vs methylated cytosines. During bisulfite conversion, methylated cytosines remain intact, and each unmethylated cytosine is converted to uracil (and eventually converted to thymine after PCR amplification). Adaptors from NEXTflex Bisulfite-Seq Barcodes (Bioo Scientific, Austin, TX, USA) were ligated to DNA fragments. Unmethylated lambda DNA was added to each DNA sample before library construction to calculate the bisulfite conversion rate.

The library from *T. dubius* individual 3060-1-4 was sequenced using Illumina HiSeq X (2 × 150 bp) at HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA. The remaining five libraries were sequenced using Illumina NovaSeq (2 × 150 bp) at the Interdisciplinary Center for Biotechnology Research, University of Florida, Gainesville, FL, USA. Detailed statistics of the libraries can be found in Supporting Information Table S1.

Analysis of DNA methylation levels

Trim_galore (v.0.5.0) was used to trim low-quality bases and remove adaptor sequences (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). bismark (v.0.22.3; Krueger & Andrews, 2011) was used to map trimmed reads to the *T. dubius* reference genome (from individual 2674-4-3-11, Oakesdale, WA, USA; Liu *et al.*, BioSample accession: SAMN36822852) with default settings. Using bismark, the alignment files were then deduplicated to eliminate the effect of excessive PCR amplification during library preparation. To eliminate the alignment efficiency bias among the three species, cytosines with a minimum coverage of three reads in all samples from all species were included in analyses below. Genome-wide weighted DNA methylation levels were calculated at different cytosine contexts by dividing the total number of methylated reads by the total number of both methylated and unmethylated reads (Schultz *et al.*, 2012). Following arcsine square root data transformation, ANOVA and *post hoc* Tukey analyses were employed to determine whether genome-wide weighted DNA methylation levels were significantly different between species. The methylation level for *T. miscellus* was compared with the mid-parent value (MPV; the average of *T. dubius* and *T. pratensis*) at different cytosine contexts using a one-sample *t*-test.

Metaplot analysis of gene regions

Metaplots of DNA methylation levels were generated using 30 325 protein-coding genes annotated from the *T. dubius* reference genome (Liu *et al.*, BioSample accession: SAMN36822852). For each gene, we calculated the weighted DNA methylation levels at three regions: 1000 bp upstream of a gene, the gene body (only includes cytosines within the coding sequence), and 1000 bp downstream of a gene. Each genomic region was divided into 20 subregions; the weighted DNA methylation level was calculated in each subregion, and the results were combined to generate the gene body methylation metaplot. The methylation levels were compared among the three species within each of the three regions (i.e. upstream, gene body, and downstream) using one-way ANOVA and *post hoc* Tukey analyses (following arcsine square root data transformation). The methylation level of *T. miscellus* was compared with the MPV using a one-sample *t*-test.

TE annotation and metaplot analysis of TE regions

Repeat annotation of the *T. dubius* reference genome was performed using REPEATMODELER (v.2.0; Flynn *et al.*, 2020) and REPEATMASKER (v.4.0.9; Smit *et al.*, 2013–2015). Briefly, REPEATMODELER was used for *de novo* identification of TEs and other repeat classes within the *T. dubius* genome. The repeat library produced by REPEATMODELER was used as input for REPEATMASKER to annotate and mask specific repetitive elements within the genome. After masking and annotation with the *de novo* library, the genome was again annotated with REPEATMASKER using internal library sequences from Repbase and DFAM (described in Smit *et al.*, 2013–2015) that were either specific to Asteraceae or general to lineages containing Asteraceae. Both annotations were combined before use in downstream analyses.

For each TE, weighted DNA methylation levels were calculated in the following three regions: 2000 bp upstream of the TE, the TE body, and 2000 bp downstream of the TE. Each region was divided into 20 subregions, and the weighted DNA methylation level was calculated for each subregion; these individual results were then combined to generate the metaplot for TE methylation. Following arcsine square root data transformation, one-way ANOVA and *post hoc* Tukey tests were used to compare the methylation levels among species within each region (i.e. upstream, TE, and downstream). A one-sample *t*-test was conducted to compare the methylation level of *T. miscellus* to the MPV. Methylation metaplots were generated using all TEs. In addition, metaplot analysis was conducted for each major TE type, that is Copia, Gypsy, long interspersed nuclear elements (LINE), and DNA transposons.

Identification of SNPs between *T. dubius* and *T. pratensis*

Genomic DNA of *T. pratensis* (Garfield, WA, USA; 3058-4-3) was extracted at the Florida Museum of Natural History (Gainesville, FL, USA), and gDNA library preparation and sequencing were performed at BGI (Shenzhen, China). Trimmed *T. pratensis* genome sequencing reads were mapped to the *T. dubius* reference genome using BWA (v.0.7.17; Li & Durbin, 2009). MarkDuplicates from PICARD (v.2.21.2) (http://broadinstitute.github.io/picard) was used to identify duplicate reads. HaplotypeCaller from GATK (v.4.1.8.1; Van der Auwera *et al.*, 2013) was used to call variants between *T. dubius* and *T. pratensis*. The identified single-nucleotide polymorphisms (SNPs) were then filtered using VariantFiltration from GATK (Van der Auwera *et al.*, 2013) with default parameters.

Identifying the subgenomic origin of *T. miscellus* reads

SNPSPLIT (v.0.5.0; Krueger & Andrews, 2016) was used to differentiate *T. dubius*- and *T. pratensis*-derived reads in *T. miscellus*. To avoid incorrect methylation calling, SNPSPLIT used SNP positions only for allele-sorting (but not identifying methylation status; Krueger & Andrews, 2016). First, the *T. dubius* reference genome was masked by SNPs identified between the two diploid species using BEDTOOLS (v.2.30.0; Quinlan & Hall, 2010). Second, trimmed *T. miscellus* reads were mapped to the masked *T. dubius* reference genome using BISMARK; the mapped BAM files were then deduplicated. Third, the subgenome-origin of *T. miscellus* reads was identified as corresponding to either *T. dubius* or *T. pratensis* by running SNPSPLIT with default parameters.

Identification of differentially methylated regions and differentially methylated genes and GO enrichment analyses of DMGs

Using METHYLKIT (v.1.24.0; Akalin *et al.*, 2012), DMRs were identified between *T. dubius* and *T. pratensis*, and between the two *T. miscellus* subgenomes (two replicates of each species/subgenome). Methylation calls were read from sorted BISMARK alignment files. Only bases having coverage of three or more reads across all samples (from both diploids and the two allotetraploid subgenomes) were included in the DMR analysis. The size of the tiling window of a DMR was 300 bp, and a DMR covered at least 10 bases with read coverage. For CG-DMRs, regions with absolute percent methylation differences larger than 35% and an adjusted *P*-value (incorporating multiple testing correction) < 0.01 were considered differentially methylated. The cutoff of methylation difference for CHG-DMRs and CHH-DMRs was 25% and 10%, respectively (adjusted *P*-value < 0.01). Genes overlapping with DMRs were defined as differentially methylated genes (DMGs). Gene Ontology (GO) enrichment analyses of DMGs were performed using the GOseq pipeline included in TRINITY (v.r20180213-2.6.5; adjusted *P*-value < 0.01; Young *et al.*, 2010).

Results

Genome-wide methylation profiles in *Tragopogon*

To understand the impacts of recent polyploidization on DNA methylation, we profiled DNA methylomes from two replicates of the diploid parents and their allopolyploid derivative using WGBS. The average depth at cytosine sites for *T. dubius*, *T. pratensis*, and *T. miscellus* was 22.1×, 14.8×, and 19.6×, respectively (Table S1). The bisulfite conversion rate was over 99.6% in all samples (Table S1). To eliminate the alignment efficiency bias between species, only shared cytosine sites with a minimum coverage of three reads across all samples were used to calculate weighted methylation levels (Table S2). The genome-wide weighted CG methylation levels based on averaging the two replicates of *T. dubius*, *T. pratensis*, and *T. miscellus* were 89.7%, 85.2%, and 86.8%, respectively (Fig. 1a). ANOVA analysis indicated that the genome-wide weighted CG methylation level in *T. dubius* was significantly different from that in *T. pratensis* (Fig. 1a; Table S3). In addition, there was no significant difference between the genome-wide weighted CG methylation level in *T. miscellus* and the MPV (Table S3). For CHG methylation, levels of 73.2%, 68.4%, and 68.5% were observed in *T. dubius*, *T. pratensis*, and *T. miscellus*, respectively (Fig. 1a). The genome-wide weighted CHG methylation level in *T. dubius* was significantly different from that in both *T. pratensis* and *T. miscellus* (Fig. 1a; Table S3), and the methylation level of *T. miscellus* was

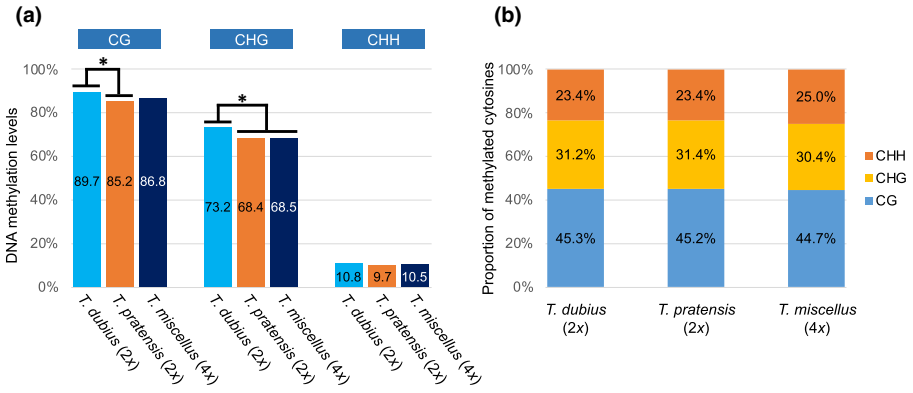

Fig. 1 Genome-wide methylation profiles for *Tragopogon dubius*, *Tragopogon pratensis*, and *Tragopogon miscellus*. (a) Genome-wide weighted DNA methylation levels. (b) Proportions of methylated cytosines in different contexts. *, *P*-value < 0.05 by ANOVA and *post hoc* Tukey analyses.

not significantly different from the MPV in the CHG context (Table S3). For CHH methylation, genome-wide weighted methylation levels in *T. dubius*, *T. pratensis*, and *T. miscellus* were 10.8%, 9.7%, and 10.5%, respectively (Fig. 1a). There was no significant difference among the three species in genome-wide weighted CHH methylation levels; *T. miscellus* was not significantly different from the MPV (Table S3). In summary, in CG and CHG cytosine contexts, the genome-wide weighted methylation levels in *T. dubius* were significantly different from those in *T. pratensis*; in all cytosine contexts, the methylation level of *T. miscellus* was not significantly different from the MPV.

The contributions of cytosine methylation from different contexts to the total number of methylated cytosines were also examined (Fig. 1b). In *T. dubius*, the proportions of methylated cytosines in CG, CHG, and CHH contexts were 45.3%, 31.2%, and 23.4%, respectively. In *T. pratensis*, 45.2%, 31.4%, and 23.4% of methylated cytosines were found in CG, CHG, and CHH, respectively. In *T. miscellus*, the proportions of methylated cytosines in CG, CHG, and CHH were 44.7%, 30.4%, and 25.0%, respectively. There was no significant difference in the proportion of methylated cytosines among the three species in any of the contexts (Table S4).

Methylation levels in gene bodies and their flanking regions

Using 30 325 protein-coding genes (Liu *et al.*, BioSample accession: SAMN36822852), methylation levels were analyzed within gene bodies and their flanking regions (1000 bp upstream and downstream of the gene) across all cytosine contexts (Fig. 2a). As reported in other flowering plants (Feng *et al.*, 2010), depletion of CG methylation around the transcription start site (TSS) and the transcription termination site (TTS) was found in all *Tragopogon* species. The CHG and CHH methylation levels within the gene body were lower than those in both flanking regions in both the diploids and the polyploid (Fig. 2a).

Statistical analyses showed that the methylation levels were significantly different between *T. dubius* and *T. pratensis* only at the gene promoter region in the CHG and CHH contexts (Fig. 2a; Table S5). *Tragopogon miscellus* was not significantly different from the MPV at either the gene body or the flanking regions in any cytosine context (Table S5).

TE annotation and methylation levels in TEs and their flanking regions

In total, we examined 534 131 annotated TEs (Table S6). Retrotransposons and DNA transposons accounted for 41.0% and 2.0% of the *T. dubius* reference genome, respectively. Of the retrotransposons, long terminal repeat (LTR) elements comprised 39.8% of the genome, with Copia and Gypsy elements accounting for 21.8% and 11.1% of the genome, respectively, and LINEs constituted 1.3%. Detailed statistics of different TE types are shown in Table S6.

For methylation metaplot analyses, we first generated metaplots using all annotated TEs (Fig. 2a). In all species, TE bodies were hypermethylated relative to the flanking regions in the CG and CHG contexts, but not in the CHH context (Fig. 2a; Table S7). In the CG and CHG contexts, the methylation levels were significantly different between *T. dubius* and *T. pratensis* at the upstream, TE, and downstream regions; in both contexts, the methylation levels in *T. miscellus* were significantly lower than the MPV at the TE body (Fig. 2a; Table S5). In the CHH context, no significant difference was identified among the three species at the upstream, TE, or downstream regions (Fig. 2b; Table S5).

Second, we generated the methylation metaplots for each major TE type, including Copia, Gypsy, and LINE elements and DNA transposons (Fig. 2b). In the CG and CHG contexts, for Copia elements, LINEs, and DNA transposons, methylation levels within TEs were significantly higher than those of the flanking regions in all three species, but these patterns were not found for Gypsy elements (Table S7). In the CHH context, elevated methylation within TEs was only found for DNA transposons and only in *T. miscellus* (Table S7).

The methylation levels in *T. dubius* and *T. pratensis* were significantly different in the CG and CHG contexts (Table S8). For Copia and Gypsy elements, in both contexts, the methylation levels in *T. dubius* were significantly different from those of *T. pratensis* in the upstream, TE, and downstream regions. For LINEs and DNA transposons, in the CG context, the two diploids were significantly different only at the body region. The CHG methylation levels in the diploids were not significantly different for LINEs (in the upstream region).

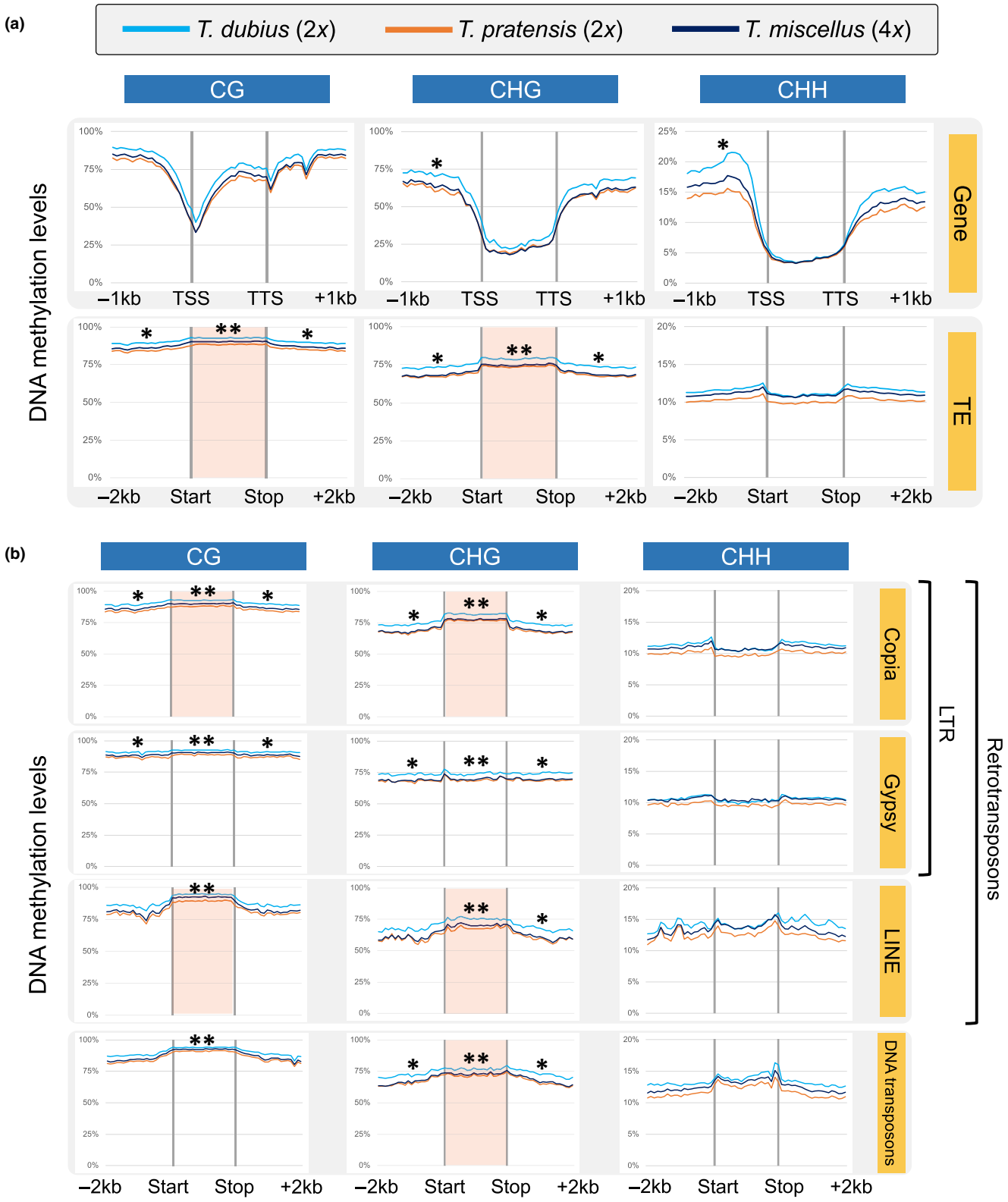

Fig. 2 Methylation metaplots at gene bodies and transposable elements (TEs) and their flanking regions. (a) Methylation metaplots generated using all genes and all TEs. (b) Methylation metaplots of different types of TEs. Regions at which the *Tragopogon miscellus* methylation level is significantly different from the mid-parent value (MPV; *P*-value < 0.05; one-sample *t*-test) are shaded in pink. TSS, transcription start site; TTS, transcription termination site. *, *P*-value < 0.05 (ANOVA test); **, *P*-value < 0.01 (ANOVA test).

The methylation level in *T. miscellus* was significantly different from the MPV only at TE bodies and only in the CG and CHG contexts (Fig. 2b; Table S8). In the CG context, the methylation level in *T. miscellus* was significantly lower and higher than the MPV for Copia elements and LINEs, respectively. In the CHG context, the methylation level in *T. miscellus* was significantly lower than the MPV for Copia elements, LINEs, and DNA transposons. For Gypsy elements, there was no significant difference between the *T. miscellus* methylation level and the MPV in any cytosine context.

SNP identification between diploids and determination of the parental origin of *T. miscellus* reads

Tragopogon pratensis reads were aligned to the *T. dubius* reference genome, and the mean and median coverage were 43.2× and 21.0×, respectively (Table S9). After variant calling and filtration, 13 218 377 SNPs were identified between *T. dubius* and *T. pratensis*. Only homozygous alternative alleles (2100 512 SNPs) were used in the reads-origin-determination study. The mapping efficiencies of *T. miscellus* individuals 3059-7-7 and 3059-21-5 to the masked *T. dubius* reference genome were 32.9% and 31.8%, respectively. After deduplication, 72.2% and 81.4% of the reads remained in individuals 3059-7-7 and 3059-21-5, respectively. In *T. miscellus* 3059-7-7, 82.8% of the reads did not differentiate between *T. dubius* and *T. pratensis*; 9.5% and 7.3% of the reads were assigned to the *T. dubius* and *T. pratensis* subgenomes, respectively, and 0.4% of the reads were classified as 'conflicting' because they contained SNPs diagnostic for both genomes. In *T. miscellus* 3059-21-5, 9.5% and 7.6% of the reads were assigned to the *T. dubius* and *T. pratensis* subgenomes, respectively; 82.6% of the reads did not differentiate between the two subgenomes, and 0.4% reads were 'conflicting'.

Identification of DMRs between the diploid parents and the two subgenomes in *T. miscellus*

Between *T. dubius* and *T. pratensis*, 3549 CG-DMRs were identified (Fig. 3). Of these, 2311 CG-DMRs (65.1%) showed hypermethylation (i.e. higher methylation level) in *T. dubius* relative to *T. pratensis*, and 1238 CG-DMRs (34.9%) showed hypomethylation (i.e. lower methylation level) in *T. dubius* relative to *T. pratensis*. Between the two subgenomes in *T. miscellus*, there were 3262 CG-DMRs: 1814 (55.6%) and 1448 (44.4%) CG-DMRs showed hypermethylation and hypomethylation, respectively, in the *T. dubius*-derived subgenome relative to the *T. pratensis*-derived subgenome. Of the CG-DMRs identified between the diploids, 64.8% (2300) of them showed parental legacy in *T. miscellus*: These regions were also differentially methylated (in the same direction) between the two *T. miscellus* subgenomes (Fig. 3). In addition, there were 1189 diploid-specific CG-DMRs that were only differentially methylated between *T. dubius* and *T. pratensis*; 902 CG-DMRs were polyploid-specific and were differentially methylated only between the two subgenomes in *T. miscellus* (Fig. 3). Sixty CG-DMRs showed one direction (e.g. hypermethylation) between the two diploids, but the other direction (e.g. hypomethylation) between the two tetraploid subgenomes.

In the CHG context, 4886 DMRs were identified between the diploids: 3341 (68.4%) and 1545 (31.6%) showed hypermethylation and hypomethylation in *T. dubius* relative to *T. pratensis*, respectively (Fig. 3). There were 4041 CHG-DMRs between the two subgenomes: 2052 (50.8%) showed hypermethylation in the *T. dubius*-derived subgenome relative to the *T. pratensis*-derived subgenome, and 1989 (49.2%) showed hypermethylation in the *T. pratensis*-derived subgenome relative to the *T. dubius*-derived subgenome (Fig. 3). In addition, 2838 (58.1%) CHG-DMRs identified between the diploids showed parental legacy in *T. miscellus*; the numbers of diploid-specific and polyploid-specific DMRs were 1892 and 1047, respectively (Fig. 3). A total of 156 CHG-DMRs showed one direction between the two diploids, but the other direction between the two tetraploid subgenomes.

In the CHH context, 18 143 DMRs were found in the diploids: 12 655 (69.8%) and 5488 (30.2%) were hypermethylated and hypomethylated in *T. dubius* relative to *T. pratensis*, respectively (Fig. 3). In *T. miscellus*, 14 535 CHH-DMRs were identified between the two subgenomes: 7405 (50.9%) and 7130 (49.1%) showed hypermethylation and hypomethylation, respectively, in the *T. dubius*-derived subgenome relative to the *T. pratensis*-derived subgenome. Of the CHH-DMRs identified between the diploids, 7440 (41.0%) showed parental legacy in *T. miscellus*; 10 467 CHH-DMRs were diploid-specific (Fig. 3). Of the CHH-DMRs identified in *T. miscellus*, 6859 were polyploid-specific. In addition, 236 CHH-DMRs showed one direction between the two diploids, but the other direction between the two tetraploid subgenomes.

In addition to the qualitative description of DMRs showing parental legacy (as mentioned in the previous section), we quantitatively examined these DMRs in each cytosine context to answer the question: How does the difference in methylation level between the diploids compare to that between the two subgenomes in the polyploid? For each DMR showing parental legacy, the difference in methylation level between the two subgenomes (represented by A) was subtracted from the difference between the diploid parents (B); the resulting absolute value (|B-A|) was used to construct the density distribution plot (Fig. S1). The mean values of |B-A| were 12.5%, 12.7%, and 5.0% in CG, CHG, and CHH contexts, respectively. The median values for |B-A| were 7.6%, 7.6%, and 3.4% in CG, CHG, and CHH contexts, respectively. Additionally, in CG, CHG, and CHH contexts, 8.2%, 16.4%, and 14.2% of the DMRs showing parental legacy exhibited substantial alteration in methylation level differences following polyploidy, respectively (Fig. S1). That is, |B-A| was greater than the cutoff defining DMR (i.e. 35%, 25%, and 10% in CG, CHG, and CHH contexts, respectively; see the Materials and Methods section).

DMG identification and GO enrichment analysis

We identified DMGs (genes overlapping with DMRs) and inferred the enriched GO terms under each DMG category (Fig. 4). In the CG context, 1580 DMGs were identified between

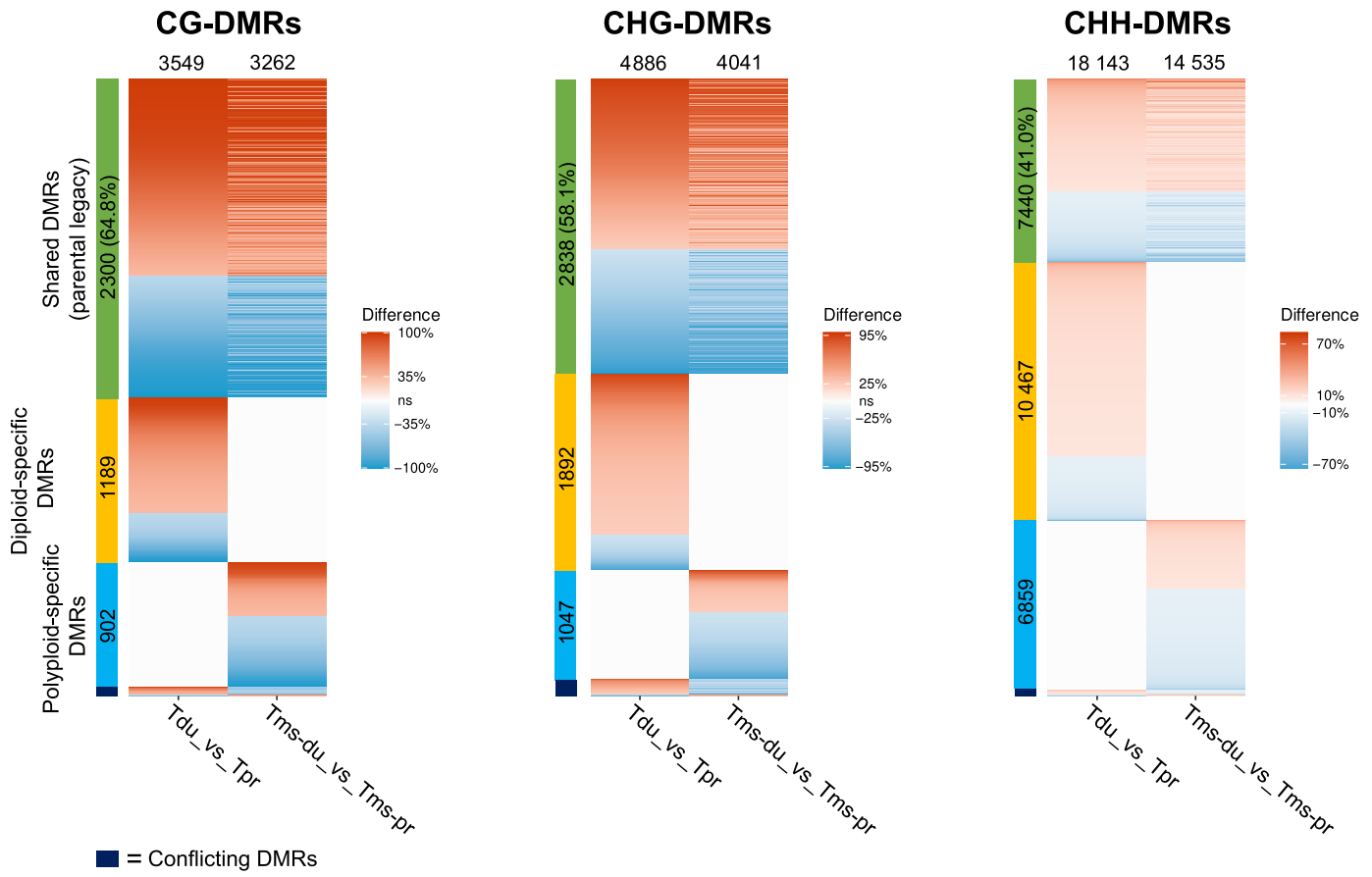

Fig. 3 Heatmap of differentially methylated regions (DMRs) identified between diploids and between the two subgenomes in *Tragopogon miscellus* in CG, CHG, and CHH contexts. Each row represents a DMR. Shared DMRs (i.e. DMRs showing parental legacy) are DMRs between *Tragopogon dubius* and *Tragopogon pratensis* that also occur between the two *T. miscellus* subgenomes. The percentage of DMRs exhibiting parental legacy was calculated by dividing the number of shared DMRs by the total number of DMRs between the diploids in each cytosine context. Diploid-specific DMRs are regions that are only differentially methylated between the diploids. Polyploid-specific DMRs represent regions that are only differentially methylated between the two subgenomes in *T. miscellus*. Conflicting DMRs are regions that show one direction (e.g. hypermethylation) between the two diploids, but the other direction (e.g. hypomethylation) between the two tetraploid subgenomes. The number of DMRs falling into each category is also shown. In CG, CHG, and CHH contexts, there were 60, 156, and 236 conflicting DMRs, respectively. A positive value in the 'Difference' legend indicates hypermethylation in either *T. dubius* or the *T. dubius-derived* subgenome in *T. miscellus* relative to the *T. pratensis* counterpart; a negative value means *T. dubius*/*T. dubius-derived* subgenome hypomethylation. Tdu_vs_Tpr: (the methylation level in *T. dubius*) – (the methylation level in *T. pratensis*); Tms-du_vs_Tms-pr: (the methylation level of the *T. dubius-derived* subgenome in *T. miscellus*) – (the methylation level of the *T. pratensis-derived* subgenome in *T. miscellus*); ns: not significantly different. Within each DMR category, DMRs are listed (from top to bottom) in descending order of the Tdu_vs_Tpr value.

the diploids (Table S10); enriched GO terms in these DMGs included membrane component and protein phosphorylation. There were 1556 CG-DMGs between the two subgenomes in the polyploid (Table S10); enriched GO terms were membrane component and glucose transporter. CG-DMGs showing parental legacy were enriched in GO terms related to membrane component. No GO term was enriched in diploid-specific CG-DMGs. RNA modification and endonuclease activity were the enriched GO terms in polyploid-specific DMGs.

In the CHG context, 1858 DMGs were identified between the diploids (Table S10); ATP binding was the only enriched GO term in these DMGs (Fig. 4). Between the two subgenomes in *T. miscellus*, ATP binding was the enriched GO term in the 1699 CHG-DMGs. CHG-DMGs showing parental legacy were enriched in ATP binding as well. We did not find any enriched GO term in the diploid-specific and polyploid-specific CHG-DMGs.

In the CHH context, 2247 DMGs between the diploids showed enriched functions related to membrane component, carbohydrate metabolic process, and hydrolase activity; 1890 DMGs between the two subgenomes were enriched in GO terms of membrane component and signal transduction (Fig. 4; Table S10). Membrane component and signal transduction were also the two GO terms enriched in CHH-DMGs showing parental legacy. Enriched GO terms in diploid-specific DMGs were membrane component and hydrolase activity. No enriched GO term was found in polyploid-specific DMGs.

Cellular component
Biological process
Molecular function

DMG Category	Enriched GO terms		
	CG	CHG	CHH
Tdu vs Tpr	Integral component of membrane Plasma membrane Protein phosphorylation	ATP binding	Plasma membrane Integral component of membrane Carbohydrate metabolic process Hydrolase activity, acting on ester bonds
Tms_du vs Tms_pr	Integral component of membrane Glucose import Plasma membrane Carbohydrate:proton symporter activity Glucose transmembrane transporter activity	ATP binding	Signal transduction Integral component of membrane
Parental legacy	Integral component of membrane Plasma membrane	ATP binding	Plasma membrane Integral component of membrane Signal transduction
Diploid-specific	na	na	Plasma membrane Integral component of membrane Hydrolase activity, acting on ester bonds
Polyploid-specific	RNA modification Endonuclease activity	na	na

Fig. 4 Gene Ontology (GO) enrichment analysis of differentially methylated genes (DMGs). All enriched GO terms had an adjusted *P*-value < 0.01. 'Tdu vs Tpr' indicates DMGs between *Tragopogon dubius* and *Tragopogon pratensis*. 'Tms_du vs Tms_pr' indicates DMGs between the *T. dubius-derived* subgenome and the *T. pratensis-derived* subgenome in *Tragopogon miscellus*. DMGs showing parental legacy are DMGs between *T. dubius* and *T. pratensis* that also occur between the two *T. miscellus* subgenomes. Diploid-specific DMGs are genes that are only differentially methylated between the diploids. Polyploid-specific DMGs represent genes that are only differentially methylated between the two subgenomes in *T. miscellus*. na means no GO term was significantly enriched.

Discussion

Genome-wide methylation profiles in *Tragopogon*

The genome-wide weighted DNA methylation levels in *Tragopogon* (means for the three *Tragopogon* species of 87.2%, 70.0%, and 10.3% for CG, CHG, and CHH methylation, respectively) are considerably higher than values for most other angiosperms examined to date in all cytosine contexts. Niederhuth *et al.* (2016) examined DNA methylation in 34 angiosperm genera and found extensive variation in genome-wide methylation levels: CG, CHG, and CHH methylation levels ranged from 30.5% to 92.5% (median of 52.0%), 9.3% to 81.2% (median of 27.0%), and 1.1% to 18.9% (median of 4.8%), respectively.

A positive correlation between genome size and genome-wide CG and CHG methylation levels was found across 34 angiosperm genera (Niederhuth *et al.*, 2016). This correlation might result from the fact that large genomes often have more TEs and DNA repeats than smaller genomes, and these TEs and DNA repeats are usually heavily methylated (Niederhuth *et al.*, 2016). The estimated genome sizes of *T. dubius* and *T. pratensis* are 2.88 and 2.71 Gb, respectively (Pires *et al.*, 2004), and TEs and DNA repeats constitute *c.* 75% of their genomes (Spoelhof *et al.*, unpublished). Therefore, the high methylation rates found in *Tragopogon* could reflect its relatively large genome size and high TE and repeat content.

Additionally, in both diploid and polyploid *Tragopogon*, *c.* 45% of the methylated cytosines were in the CG context, followed by CHG and CHH (Fig. 1b). CG methylation is the predominant type of methylation in both plants and animals (Niederhuth *et al.*, 2016; Schmitz *et al.*, 2019; De Mendoza *et al.*, 2020).

Genome-wide and locus-specific DNA methylation dynamics in polyploids

In the allotetraploid *Tragopogon miscellus*, the genome-wide weighted DNA methylation levels were not significantly different from the MPV in any cytosine context, indicating an overall additive methylation landscape in the polyploid (Fig. 1a; Table S3). The identification of diploid-specific and polyploid-specific DMRs suggests that regions showing hypermethylation following WGD counterbalanced those showing hypomethylation, leading to the additive genome-wide methylation level in *T. miscellus* (Fig. 3). In addition, for DMRs showing parental legacy, the difference in methylation level changed substantially following polyploidy in some regions (Fig. S1); these regions, along with diploid-specific and polyploidy-specific DMRs,

illustrate the dynamic methylation alterations that occur following WGD in *Tragopogon*.

Genome-wide and locus-specific DNA methylation dynamics have been reported in a few polyploids, mostly in synthetics. In rice, although genome-wide methylation levels of synthetic allotetraploids (derived from a cross between *Oryza sativa* ssp. *japonica* and *O. sativa* ssp. *indica*) were similar to those of the diploid F_1 hybrids and the parents, DNA methylation changes in the tetraploids displayed a locus-specific pattern, with loci showing hyper- and hypomethylation largely offsetting each other (Li *et al.*, 2019). Specifically, 55.8%, 48.2%, and 33.5% of DMRs showed parental legacy following WGD in CG, CHG, and CHH contexts, respectively (the calculation is based on the data presented in fig. 4a of Li *et al.*, 2019). Novel DMRs were also found in these synthetic polyploids (Li *et al.*, 2019). In *T. miscellus*, 64.8% of CG-DMRs, 58.1% of CHG-DMRs, and 41.0% of CHH-DMRs exhibited parental legacy (Fig. 3). In both synthetic polyploid rice and natural allotetraploid *T. miscellus*, a decreasing percentage of DMRs exhibits parental legacy across the CG to CHG to CHH contexts. Using WGBS, the genome-wide methylation profiles were analyzed in synthetic *Brassica napus* allotetraploids and their progenitor lines: the methylation differences between the two subgenomes resulted from a combination of parental legacy and methylation changes following hybridization and polyploidization (Bird *et al.*, 2021). In *Senecio* (Asteraceae), *c.* 13.4% of loci showed nonadditive DNA methylation changes in synthetic triploid F_1 hybrids (*S.* × *baxteri*; Hegarty *et al.*, 2011). Multiple studies investigating DNA methylation dynamics in recently formed natural polyploids have also been reported. For example, nonadditive methylation changes were observed in natural allopolyploid *Spartina* (Salmon *et al.*, 2005). In both *Spartina* hybrids (*S.* × *neyrautii* and *S.* × *townsendii*) and the allopolyploid *S. anglica*, MSAP analyses showed that 30% of the parental DNA methylation patterns were altered, and most of the DNA methylation changes resulted from hybridization rather than genome doubling *per se* (Salmon *et al.*, 2005).

Dynamic TE methylation changes following polyploidy

Polyploidy-induced epigenetic changes appear to broadly and selectively impact TE regions in a genome, and the response to WGD (either hypomethylation or hypermethylation) varies among different species and TE types (Parisod *et al.*, 2010; Vicient & Casacuberta, 2017). The methylation level in *T. miscellus* is largely additive of the diploid parents, especially in gene regions (Figs 1, 2a). However, in CG and CHG contexts, *T. miscellus* methylation levels at TEs were significantly lower than the MPV (Fig. 2a). Hypomethylation of TEs was observed in other newly formed allopolyploids (Parisod *et al.*, 2009; Edger *et al.*, 2017). Based on methyl-sensitive transposon display (a TE-targeted MSAP approach), the CG methylation level in regions flanking TE insertions in natural allopolyploid *Spartina anglica* was lower than that of the diploid parents (Parisod *et al.*, 2009). In addition, significantly more DNA methylation changes were found in the vicinity of TEs compared with random genomic loci (Parisod *et al.*, 2009). Similarly, in *Tragopogon*, nonadditive methylation changes following polyploidy were only identified at TE regions (Fig. 2). In natural allopolyploid *Mimulus peregrinus*, at TEs and their flanking regions, the CHH methylation level of the *M. luteus*-derived subgenome was lower than that of the parent *M. luteus* (Edger *et al.*, 2017). By contrast, in synthetic *Brassica napus* allotetraploids, transgressive CHH hypermethylation (i.e. surpassing parental methylation levels) was found at LTR retrotransposons (Bird *et al.*, 2021). Similarly, synthetic autotetraploid rice derived from diploid *O. sativa* showed hypermethylation in DNA transposons compared with the diploid. These results indicate that the direction of methylation changes following polyploidy at TEs varies across species, further demonstrating the complexity of the impact of WGD on DNA methylation dynamics.

The impact of WGD on methylation dynamics varies among different TE types in *Tragopogon* (Fig. 2b). The CG methylation level in *T. miscellus* was significantly lower than the MPV for Copia elements, but significantly higher than the MPV for LINEs; in the CHG context, methylation in *T. miscellus* was significantly lower than the MPV for Copia elements, LINEs, and DNA transposons. However, the methylation levels in *T. miscellus* were additive for Gypsy elements in both CG and CHG contexts. The chromosome localization of TEs may affect their methylation dynamics following polyploidy (Vicient & Casacuberta, 2017). Gypsy elements are preferentially located in heterochromatin and away from genes, and Copia elements and DNA transposons are prone to insert in euchromatin and near genes (Galindo-González *et al.*, 2017; Vicient & Casacuberta, 2017). Indeed, in the *T. dubius* reference genome, Gypsy elements were significantly farther away from genes than other types of TEs (Fig. S2). In *T. miscellus*, the additive methylation levels for Gypsy elements may help to maintain the heterochromatic state of the associated regions, which is essential for genome stability (Janssen *et al.*, 2018) and the proper functioning of these regions (Dernburg *et al.*, 1996; Vicient & Casacuberta, 2017).

Distinct methylation profiles across various types of TEs have also been observed in other polyploids (Vicient & Casacuberta, 2017). Synthetic autotetraploid rice derived from diploid *O. sativa* showed widespread hypermethylation for DNA transposons but hypomethylation for retrotransposons (including Copia, Gypsy, and LINE categories) in CG and CHH contexts (Zhang *et al.*, 2015). Compared with the diploid, synthetic autotetraploid water spinach (*Ipomoea aquatica*) exhibited hypermethylation for Copia, Gypsy, and LINE retrotransposons in CHG and CHH contexts; DNA transposons showed similar methylation levels between the diploid and the tetraploid (Hao *et al.*, 2023). In synthetic allopolyploid rice, the methylation levels of both retrotransposons and DNA transposons were significantly different from the *in silico* hybrid in all cytosine contexts (Li *et al.*, 2019). However, detailed comparisons of the impact of WGD on DNA methylation across various TE types (e.g. within retrotransposons) are lacking for synthetic allopolyploid rice (Li *et al.*, 2019) and other allopolyploids, further emphasizing the contribution of our current work on allopolyploid *Tragopogon*.

Because TEs are usually heavily methylated and silenced (Law & Jacobsen, 2010; Ito & Kakutani, 2014), the reduced CG and CHG methylation levels compared with MPV levels found in our work may lead to TE activation in *T. miscellus*. Indeed, polyploidization frequently triggers TE activation (Vicient & Casacuberta, 2017; Nieto Feliner *et al.*, 2020). In addition, TE dynamics following WGD have effects on genome size, gene function, and regulation of expression and may contribute to the adaptation of polyploids to environmental conditions (Vicient & Casacuberta, 2017; Nieto Feliner *et al.*, 2020). Nevertheless, because the genome sequence of *T. miscellus* is not yet available, hindering a direct comparison between diploid and polyploid genomes, additional assessment of the impact of WGD on TE activity (i.e. transcription, mobilization, and copy number) will require further investigation. An unanswered question is how WGD-induced nonadditive methylation changes at TEs affect their expression and transposition in *T. miscellus*.

Future studies should integrate genome-wide research on DNA methylation and regulation of gene expression in *Tragopogon* polyploids and examine how polyploidy-induced methylation changes impact homeolog expression bias (i.e. unequal expression levels of the two homeologs of a gene in the allopolyploid) in *T. miscellus*. DNA methylation may affect gene expression through the regulation of TEs that are adjacent to genic regions. Examination of DNA methylation patterns across genome assemblies of *T. miscellus* and its diploid parents may reveal whether DNA methylation of TEs could explain biparental homeolog expression bias observed in *T. miscellus* (e.g. Boatwright *et al.*, 2018). Likewise, mapping methylation changes onto chromosome-level assemblies may reveal linkages between methylation dynamics and karyotypic variants as observed in *T. miscellus* (Lim *et al.*, 2008; Chester *et al.*, 2012; Spoelhof *et al.*, 2017). For example, Chester *et al.* (2012) found chromosomal rearrangements (translocations) in *T. miscellus*; could some of the methylation changes observed here activate TEs, by de-methylating them, resulting in TE transposition and genome reorganization, which in turn could cause the observed translocations? Similarly, the compensated aneuploidy observed in *T. miscellus* (Chester *et al.*, 2012) could explain some of the observed examples of nonparental legacy. That is, a 'hypermethylated' observation (a higher methylation level in *T. dubius* than in *T. pratensis*) might appear 'hypomethylated' in *T. miscellus* because there are no chromosomes of *T. dubius* carrying that locus due to compensated aneuploidy. Finally, in this study, the impacts of hybridization and polyploidization itself on DNA methylation cannot be disentangled; thus, it is unclear whether novel methylation patterns emerge due to hybridization, polyploidization, or both. Comparison among F_1 hybrids, newly formed synthetic polyploids, and naturally occurring polyploids in *Tragopogon* could inform this question.

In summary, using a high-throughput WGBS approach, the present study examined for the first time both the overall and locus-specific patterns of the DNA methylation dynamics following recent natural polyploidy. Our work revealed that the genome-wide weighted DNA methylation levels of allopolyploid *T. miscellus* were largely intermediate between those of its diploid parents and reflect parental legacy. However, nonadditive CG and CHG methylation in various types of TEs and novel DMRs identified in the polyploid also indicate dynamic and rapid epigenetic changes following recent WGD in *Tragopogon*. These epigenetic data join a growing body of evidence that allopolyploidy in *Tragopogon* generates genetically (and epigenetically) diverse individuals that vary in morphology, karyotype, gene content, gene expression (as reviewed previously), and the potential for duplicate gene regulation through DNA methylation. Sorting and recombining of these karyotypic, genetic, and epigenetic variants lead to diverse arrays of polyploid individuals within populations – and across populations of independent origin – providing the basis for adaptation and the origin of novelty.

Acknowledgements

This work was supported by the National Science Foundation grants IOS-1923234 and DEB-2043478 to DES and PSS and MCB-2242696 to RJS. The authors thank Taliesin J. Kinser and Kasey K. Pham for helpful discussion. The authors also thank Christian Parisod and an anonymous reviewer for their helpful and constructive suggestions.

Competing interests

None declared.

Author contributions

DES, RJS, PSS and SS designed the project. SS and CLE prepared the WGBS libraries. SS, MAG, JLB, JPS, LJ and XL analyzed the data. SS, DES, RJS and PSS wrote the manuscript. All authors reviewed and approved the manuscript.

ORCID

J. Lucas Boatwright https://orcid.org/0000-0003-4371-2457
Matthew A. Gitzendanner https://orcid.org/0000-0002-7078-4336
Lexiang Ji https://orcid.org/0000-0003-2670-8413
Robert J. Schmitz https://orcid.org/0000-0001-7538-6663
Shengchen Shan https://orcid.org/0000-0002-9472-0986
Douglas E. Soltis https://orcid.org/0000-0001-8638-4137
Pamela S. Soltis https://orcid.org/0000-0001-9310-8659
Jonathan P. Spoelhof https://orcid.org/0000-0002-9630-7095

Data availability

The raw sequencing files have been deposited in the BioProject database from NCBI (ID: PRJNA996185). All scripts are available at: https://github.com/GatorShan/Tragopogon-Methylation-Project.

References

Agius DR, Kapazoglou A, Avramidou E, Baranek M, Carneros E, Caro E, Castiglione S, Cicatelli A, Radanovic A, Ebejer JP *et al.* 2023. Exploring the crop epigenome: a comparison of DNA methylation profiling techniques. *Frontiers in Plant Science* **14**: 1181039.

Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, Mason CE. 2012. methylkit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biology* **13**: R87.

Bird KA, Niederhuth CE, Ou S, Gehan M, Pires JC, Xiong Z, VanBuren R, Edger PP. 2021. Replaying the evolutionary tape to investigate subgenome dominance in allopolyploid *Brassica napus*. *New Phytologist* **230**: 354–371.

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

Buggs RJ, Chamala S, Wu W, Tate JA, Schnable PS, Soltis DE, Soltis PS, Barbazuk WB. 2012. Rapid, repeated, and clustered loss of duplicate genes in allopolyploid plant populations of independent origin. *Current Biology* **22**: 248–252.

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

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

Cao X, Jacobsen SE. 2002. Locus-specific control of asymmetric and CpNpG methylation by the *DRM* and *CMT3* methyltransferase genes. *Proceedings of the National Academy of Sciences, USA* **99**(Suppl 4): 16491–16498.

Chalhoub B, Denoeud F, Liu S, Parkin IA, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B *et al.* 2014. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* **345**: 950–953.

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

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

Chester M, Riley RK, Soltis PS, Soltis DE. 2015. Patterns of chromosomal variation in natural populations of the neoallotetraploid *Tragopogon mirus* (Asteraceae). *Heredity* **114**: 309–317.

Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE. 2008. Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* **452**: 215–219.

Cuerda-Gil D, Slotkin RK. 2016. Non-canonical RNA-directed DNA methylation. *Nature Plants* **2**: 16163.

De Mendoza A, Lister R, Bogdanovic O. 2020. Evolution of DNA methylome diversity in eukaryotes. *Journal of Molecular Biology* **432**: 1687–1705.

Dernburg AF, Sedat JW, Hawley RS. 1996. Direct evidence of a role for heterochromatin in meiotic chromosome segregation. *Cell* **86**: 135–146.

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**: 1–52.

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

Feng S, Cokus SJ, Zhang X, Chen PY, Bostick M, Goll MG, Hetzel J, Jain J, Strauss SH, Halpern ME *et al.* 2010. Conservation and divergence of methylation patterning in plants and animals. *Proceedings of the National Academy of Sciences, USA* **107**: 8689–8694.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences, USA* **117**: 9451–9457.

Fox D, Soltis DE, Soltis PS, Ashman TL, Van de Peer Y. 2020. Polyploidy: a driving biological force from cells to ecosystems, and from agriculture to medicine. *Trends in Cell Biology* **30**: 688–694.

Galindo-González L, Mhiri C, Deyholos MK, Grandbastien MA. 2017. LTR-retrotransposons in plants: engines of evolution. *Gene* **626**: 14–25.

Gong Z, Morales-Ruiz T, Ariza RR, Roldán-Arjona T, David L, Zhu JK. 2002. *ROS1*, a repressor of transcriptional gene silencing in *Arabidopsis*, encodes a DNA glycosylase/lyase. *Cell* **111**: 803–814.

Greaves IK, Groszmann M, Ying H, Taylor JM, Peacock WJ, Dennis ES. 2012. Trans chromosomal methylation in *Arabidopsis* hybrids. *Proceedings of the National Academy of Sciences, USA* **109**: 3570–3575.

Hao Y, Su X, Li W, Li L, Zhang Y, Mumtaz MA, Shu H, Cheng S, Zhu G, Wang Z. 2023. The creation of autotetraploid provides insights into critical features of DNA methylome changes after genome doubling in water spinach (*Ipomoea aquatica* Forsk). *Frontiers in Plant Science* **14**: 1155531.

Haudry A, Cenci A, Ravel C, Bataillon T, Brunel D, Poncet C, Hochu I, Poirier S, Santoni S, Glémin S *et al.* 2007. Grinding up wheat: a massive loss of nucleotide diversity since domestication. *Molecular Biology and Evolution* **24**: 1506–1517.

Hegarty MJ, Batstone TO, Barker GL, Edwards KJ, Abbott RJ, Hiscock SJ. 2011. Nonadditive changes to cytosine methylation as a consequence of hybridization and genome duplication in *Senecio* (Asteraceae). *Molecular Ecology* **20**: 105–113.

Ito H, Kakutani T. 2014. Control of transposable elements in *Arabidopsis thaliana*. *Chromosome Research* **22**: 217–223.

Janssen A, Colmenares SU, Karpen GH. 2018. Heterochromatin: guardian of the genome. *Annual Review of Cell and Developmental Biology* **34**: 265–288.

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

Kawakatsu T, Huang SS, Jupe F, Sasaki E, Schmitz RJ, Urich MA, Castanon R, Nery JR, Barragan C, He Y *et al.* 2016. Epigenomic diversity in a global collection of *Arabidopsis thaliana* accessions. *Cell* **166**: 492–505.

Krueger F, Andrews SR. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**: 1571–1572.

Krueger F, Andrews SR. 2016. SNPsplit: allele-specific splitting of alignments between genomes with known SNP genotypes. *F1000Research* **5**: 1479.

Kumar S, Mohapatra T. 2021. Dynamics of DNA methylation and its functions in plant growth and development. *Frontiers in Plant Science* **12**: 596236.

Landis JB, Soltis DE, Li Z, Marx HE, Barker MS, Tank DC, Soltis PS. 2018. Impact of whole-genome duplication events on diversification rates in angiosperms. *American Journal of Botany* **105**: 348–363.

Law JA, Jacobsen SE. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics* **11**: 204–220.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**: 1754–1760.

Li N, Xu C, Zhang A, Lv R, Meng X, Lin X, Gong L, Wendel JF, Liu B. 2019. DNA methylation repatterning accompanying hybridization, whole genome doubling and homoeolog exchange in nascent segmental rice allotetraploids. *New Phytologist* **223**: 979–992.

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

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR. 2008. Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133**: 523–536.

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

Nie WF. 2021. DNA methylation: from model plants to vegetable crops. *Biochemical Society Transactions* **49**: 1479–1487.

Niederhuth CE, Bewick AJ, Ji L, Alabady MS, Kim KD, Li Q, Rohr NA, Rambani A, Burke JM, Udall JA *et al.* 2016. Widespread natural variation of DNA methylation within angiosperms. *Genome Biology* **17**: 194.

Niederhuth CE, Schmitz RJ. 2017. Putting DNA methylation in context: from genomes to gene expression in plants. *Biochimica et Biophysica Acta (BBA) – Gene Regulatory Mechanisms* **1860**: 149–156.

Nieto Feliner G, Casacuberta J, Wendel JF. 2020. Genomics of evolutionary novelty in hybrids and polyploids. *Frontiers in Genetics* **11**: 792.

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

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

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

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

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.

Renny-Byfield S, Wendel JF. 2014. Doubling down on genomes: polyploidy and crop plants. *American Journal of Botany* **101**: 1711–1725.

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

Schmitz RJ, Lewis ZA, Goll MG. 2019. DNA methylation: shared and divergent features across eukaryotes. *Trends in Genetics* **35**: 818–827.

Schultz MD, Schmitz RJ, Ecker JR. 2012. 'Leveling' the playing field for analyses of single-base resolution DNA methylomes. *Trends in Genetics* **28**: 583–585.

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**: 701.

Shan S, Boatwright JL, Liu X, Chanderbali AS, Fu C, Soltis PS, Soltis DE. 2020. Transcriptome dynamics of the inflorescence in reciprocally formed allopolyploid *Tragopogon miscellus* (Asteraceae). *Frontiers in Genetics* **11**: 888.

Smit AFA, Hubley R, Green P. 2013–2015. *RepeatMasker Open-4.0.* [WWW document] URL http://www.repeatmasker.org [accessed 9 November 2023].

Soltis DE, Buggs RJA, 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: Soltis PS, Soltis DE, eds. *Polyploidy and genome evolution*. New York, NY, USA: Springer, 271–292.

Soltis DE, Mavrodiev EV, Brukhin V, Roalson EH, Albach DC, Godden GT, Alexeev YE, Gitzendanner MA, Freeman CC, Suárez-Santiago VN *et al.* 2023. *Tragopogon pratensis*: multiple introductions to North America, circumscription, and the formation of the allotetraploid *T. miscellus*. *Taxon* **72**: 848–861.

Soltis DE, Mavrodiev EV, Gitzendanner MA, Alexeev YE, Godden GT, Soltis PS. 2022. *Tragopogon dubius*: multiple introductions to North America and the formation of the New World tetraploids. *Taxon* **71**: 1287–1298.

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

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

Soltis PS, Liu X, Marchant DB, Visger CJ, Soltis DE. 2014. Polyploidy and novelty: Gottlieb's legacy. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**: 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. 1995. Genetic variation in *Tragopogon* species: additional origins of the allotetraploids *T. mirus* and *T. miscellus* (Compositae). *American Journal of Botany* **82**: 1329–1341.

Soltis PS, Soltis DE. 2009. The role of hybridization in plant speciation. *Annual Review of Plant Biology* **60**: 561–588.

Soltis PS, Soltis DE. 2016. Ancient WGD events as drivers of key innovations in angiosperms. *Current Opinion in Plant Biology* **30**: 159–165.

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**: 1484–1492.

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

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

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

Tran RK, Henikoff JG, Zilberman D, Ditt RF, Jacobsen SE, Henikoff S. 2005. DNA methylation profiling identifies CG methylation clusters in *Arabidopsis* genes. *Current Biology* **15**: 154–159.

Urich MA, Nery JR, Lister R, Schmitz RJ, Ecker JR. 2015. MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. *Nature Protocols* **10**: 475–483.

Van de Peer Y, Ashman TL, Soltis PS, Soltis DE. 2021. Polyploidy: an evolutionary and ecological force in stressful times. *Plant Cell* **33**: 11–26.

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J *et al.* 2013. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics* **43**: 11.10.1–11.10.33.

Vicient CM, Casacuberta JM. 2017. Impact of transposable elements on polyploid plant genomes. *Annals of Botany* **120**: 195–207.

Wendel JF, Lisch D, Hu G, Mason AS. 2018. The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation. *Current Opinion in Genetics & Development* **49**: 1–7.

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, USA* **106**: 13875–13879.

Young MD, Wakefield MJ, Smyth GK, Oshlack A. 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology* **11**: R14.

Zhang J, Liu Y, Xia EH, Yao QY, Liu XD, Gao LZ. 2015. Autotetraploid rice methylome analysis reveals methylation variation of transposable elements and their effects on gene expression. *Proceedings of the National Academy of Sciences, USA* **112**: E7022–E7029.

Zhang Q, Wang D, Lang Z, He L, Yang L, Zeng L, Li Y, Zhao C, Huang H, Zhang H *et al.* 2016. Methylation interactions in *Arabidopsis* hybrids require RNA-directed DNA methylation and are influenced by genetic variation. *Proceedings of the National Academy of Sciences, USA* **113**: E4248–E4256.

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

Zhang Y, Wendte JM, Ji L, Schmitz RJ. 2020. Natural variation in DNA methylation homeostasis and the emergence of epialleles. *Proceedings of the National Academy of Sciences, USA* **117**: 4874–4884.

Zilberman D. 2017. An evolutionary case for functional gene body methylation in plants and animals. *Genome Biology* **18**: 87.

Supporting Information

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

Fig. S1 Quantitative analysis of differentially methylated regions showing parental legacy in each cytosine context.

Fig. S2 The distance of DNA transposons, LINEs, Copia elements, and Gypsy elements to the closest genes in the *Tragopogon dubius* reference genome.

Table S1 Statistics of the MethylC-seq libraries.

Table S2 Statistics of the number of cytosine sites.

Table S3 Statistics of genome-wide weighted DNA methylation levels in *Tragopogon dubius*, *Tragopogon pratensis*, and *Tragopogon miscellus*.

Table S4 Statistics of the proportions of methylated cytosines from different contexts in *Tragopogon dubius*, *Tragopogon pratensis*, and *Tragopogon miscellus*.

Table S5 Statistical analysis of the DNA methylation levels among *Tragopogon dubius*, *Tragopogon pratensis*, and *Tragopogon miscellus* in gene and TE regions.

Table S6 Statistics of different types of TEs in the *Tragopogon dubius* reference genome.

Table S7 Statistical analysis of the DNA methylation levels between TEs and their flanking regions in *Tragopogon dubius*, *Tragopogon pratensis*, and *Tragopogon miscellus*.

Table S8 Statistical analysis of the DNA methylation levels among *Tragopogon dubius*, *Tragopogon pratensis*, and *Tragopogon miscellus* in various types of TE regions.

Table S9 Statistics of the *Tragopogon pratensis* library for SNP identification.

Table S10 Number of differentially methylated genes (DMGs) in the diploids and the polyploid.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.