Small RNAs regulation and genomic harmony: insights into allopolyploid evolution in marsh orchids (*Dactylorhiza*)

Mimmi C. Eriksson[1,2,3*], Matthew Thornton[1], Emiliano Trucchi[1,4], Thomas M. Wolfe[1,5], Francisco Balao[1,6], Mikael Hedrén[7], Ovidiu Paun[1*]

[1] Botany and Biodiversity Research, University of Vienna, Rennweg 14, A-1030 Vienna, Austria.

[2] Vienna Graduate School of Population Genetics, Veterinärplatz 1, A-1210 Vienna, Austria.

[3] Department of Plant Biology, Swedish University of Agricultural Sciences, SE-750 07 Uppsala, Sweden

[4] Marche Polytechnic University, Ancona, Italy

[5] Institute of Forest Entomology, Forest Pathology and Forest Protection (IFFF), BOKU, Peter-Jordan-Straße 82/I, 1190 Vienna, Austria

[6] Departamento de Biologia Vegetal y Ecologia, University of Seville, Sevilla, Spain;

[7] Department of Biology, Lund University, Sölvegatan 37, S-22362, Lund, Sweden;

**Corresponding authors: mimmi.c.eriksson@gmail.com; ovidiu.paun@univie.ac.at*

Summary

Hybridization and polyploidy are major drivers of plant diversification, often leading to novel ecological traits through changes in gene regulation. Among these, small RNAs (smRNAs) play a key role, particularly through their interactions with transposable elements (TEs). Using a common garden setup, we comparatively investigated how smRNAs, including those targeting TEs, contribute to post-allopolyploid genome stabilization and ecological divergence in two sibling allotetraploid marsh orchids (*Dactylorhiza majalis* and *D. traunsteineri*) and their diploid progenitors. Despite independent origins, both allopolyploids exhibited substantial overlap in smRNA targeting, including transgressive overexpression around genes linked to stress response, cell division and transcriptional regulation. TE-targeting 24 nt smRNAs followed patterns consistent with subgenome size and origin, while shorter 20–23 nt smRNAs displayed greater divergence. The older allotetraploid *D. majalis* exhibited enhanced smRNA-mediated fine-tuning of gene copy number regulation, while its younger sibling *D. traunsteineri* showed more pronounced non-additive expression patterns, suggesting ongoing resolution of post-polyploid meiotic and mitotic instability. These findings reveal a two-phase smRNA response to allopolyploidy: early genome stabilization followed by regulatory fine-tuning. Our study highlights the central role of smRNAs in resolving genomic conflict, enabling functional divergence, and driving ecological innovation during polyploid evolution.

Keywords

allopolyploidy, *Dactylorhiza*, gene regulation, marsh orchids, small RNAs, subgenome dominance, transposable elements

Introduction

Allopolyploidization—the combination of whole-genome duplication (WGD) and interspecific hybridization—is a potent evolutionary force, generating genetic diversity, gene redundancy, and buffering capacity that underpin evolutionary innovation and ecological diversification. However, the initial merger of divergent genomes into a single nucleus frequently destabilizes genome regulation (i.e., triggering “genomic shock”, McClintock 1984), often on the background of a severe genetic bottleneck.

Small noncoding RNAs (smRNAs) have emerged as key mediators during post-polyploidization genome stabilization (e.g., Ha et al. 2009; Shen et al. 2017; Cavé-Radet et al. 2020). These short RNA molecules supress transposable elements (TEs) and fine-tune gene expression, promoting both genome integrity and adaptive responses. Specifically, repeat-associated or heterochromatic small-interfering RNAs (ra-siRNAs/het-siRNAs; Shen et al. 2017) help maintain chromatin stability, while microRNAs (miRNAs) can modulate gene expression with phenotypic implications (Ha et al. 2009; Ghani et al. 2014; Jiang et al. 2021). In addition, plant smRNAs orchestrate abiotic and biotic stress responses (Kaur et al. 2021; Tiwari and Rajam 2022; Zhan and Meyers 2024), potentially facilitating adaptation to novel, non-parental environments. These dual roles make smRNAs integral to allopolyploid success.

In general, smRNAs operate through three main mechanisms: (1) post-transcriptional gene silencing (PTGS), binding to RNAs to repress translation or promote cleavage and degradation; (2) transcriptional gene silencing (TGS), directing epigenetic silencing at homologous genomic loci; and (3) splicing regulation, interacting with pre-mRNAs to influence splice-site selection (Borges and Martienssen 2015; Wang and Chekanova 2016). In plants, smRNAs are classified by biogenesis into miRNAs (20–22 nt long, hairpin-derived), and small-interfering RNAs (siRNAs, from double stranded RNA) (Borges and Martienssen 2015; Chen and Rechavi 2021), with siRNAs further divided into: (*i*) ra-siRNAs (24 nt and primarily involved in TGS), (*ii*) secondary siRNAs (~21–22nt long, which can function in both TGS and PTGS pathways), and (*iii*) natural antisense transcript siRNA (nat-siRNA, 21–24 nt, acting both in *cis* and *trans* via PTGS) (Borges and Martienssen 2015; Chen and Rechavi 2021).

Recent reviews (e.g., Ariel and Manavella 2021; Tossolini et al. 2025) highlight the pivotal link between TE mobilization and smRNA evolution. TEs not only contribute regulatory capacity through enhancer-derived transcripts (e.g., in wheat; TE-initiated enhancer-like elements, Xie et al. 2023), but also spur the emergence of smRNAs and long non-coding

82 RNAs that diversify gene regulation and support genome plasticity (Hassan et al. 2024;
83 Tossolini et al. 2025).

84 WGD, similarly to environmental stress, is known to lead to TE activation, resulting in
85 accumulation of repeat-derived smRNAs, boosting both TGS and PTGS pathways (Ito et al.
86 2011; Kenan-Eichler et al. 2011; Marí-Ordóñez et al. 2013; Fultz et al. 2015; Xiao and Wang
87 2025). For example, resynthesized *Arabidopsis* allotetraploids rapidly restore ra-siRNA
88 populations after WGD, while miRNAs show non-additive expression patterns and play key
89 roles in stabilizing hybrid genomes (Ha et al. 2009). Similarly, early generations *Brassica*
90 *napus* allotetraploids exhibit a surge in TE-derived 21 nt siRNAs, especially from Gypsy
91 elements, helping to silence newly active TEs (Martinez Palacios et al. 2019). Cavé-Radet et
92 al. (2020, 2023) observed stronger miRNA responses in allopolyploid *Spartina* compared to
93 homoploid hybrids, suggesting a more pronounced regulatory impact of WGD than
94 hybridization alone (Cavé-Radet et al. 2020). On the other hand, the divergence between
95 progenitors determines the extent of genomic shock in wheat allotetraploids, with enhanced
96 smRNA responses associated with TEs in hybrid crosses between parents with differing TE
97 load, compared to those with similar TE content (Jiao et al. 2018).

98 Despite growing interest in the role of smRNAs in allopolyploid evolution, most research has
99 focused on early-generation or synthetic polyploids. To our knowledge, only Cavé-Radet et
100 al. (2020) have examined smRNAs in wild allopolyploids, albeit also in early stages (*Spartina*
101 F_1 hybrids and *S. anglica*, ~170 years old). Thus, the long-term dynamics, regulatory roles
102 and potential ecological relevance of smRNAs in more established allopolyploid lineages
103 remain poorly explored.

104 In this study, we focus on two ecologically-distinct, well-established, sibling orchid
105 allotetraploids, *Dactylorhiza majalis s.str.* and *D. traunsteineri*, to address the long-term fate
106 of smRNAs following allopolyploidization. The marsh orchids within *D. majalis s.l.* includes
107 several independently formed sibling allotetraploids, each likely with multiple origins
108 involving overlapping stocks of parental populations (Brandrud et al. 2020). The focal
109 allotetraploids here share a complex history of repeated polyploidization and/or hybridization
110 and back-crossing, but are today reasonably well separated in morphology and habitat
111 preferences (Fig. 1a; Wolfe et al. 2023). They are estimated to have formed approximately
112 104k (*D. majalis*) and 74k (*D. traunsteineri*) generations ago (Hawranek 2021), from the
113 diploid progenitors *D. fuchsii* (maternal, smaller genome) and *D. incarnata* (paternal, larger
114 genome) (Fig. 1a) (Aagard et al. 2005; Brandrud et al. 2020; Eriksson et al. 2022).
115 Transcriptomic comparisons of the diploids have revealed large-scale gene expression
116 differences, importantly, including differences in smRNA-mediated regulation of TEs and

PTGS-related pathways (Balao et al. 2017). Further, Eriksson et al. (2022) reported significant differences in both genome size and TE load between the two diploid progenitors.

In short, *D. traunsteneri* prefers marshes and fens with high, whereas *D. majalis s.str.* (hereafter *D. majalis*) grows in mesic to moist meadows.

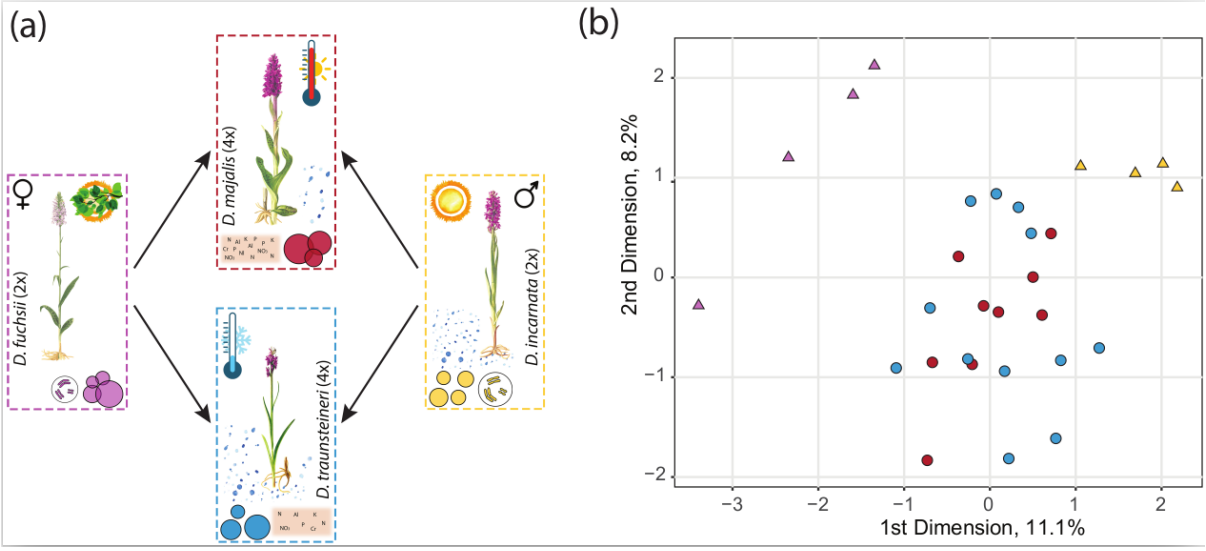


Figure 1. Two diploids (*D. fuchsii* and *D. incarnata*) and their sibling allotetraploids (*D majalis* and *D. traunsteineri*) are studied in this paper. (a) Characterization of the four species. Coloured circles and their spatial distribution towards one another reflect population densities. The cells indicate difference in genome size between the diploids and gender symbols indicate the direction of the cross to produce the allopolyploids. Other symbols represent broad scale ecological preferences. (b) Multidimensional scaling plot of smRNA count data with colours representing the species frames in (a) while triangles indicate diploids, and circles, polyploid accessions.

Distinct ecological, spatial, morphological and physiological properties have been reported for the two sibling allopolyploids (Paun et al. 2010, 2011; Balao et al. 2016; Wolfe et al. 2023). *D. majalis* is distributed across continental Europe, ranging from the Pyrenees to Scandinavia, with a broader ecological niche (i.e., mesic to moist meadows) and less geographically-driven genetic differentiation than *D. traunsteineri* which is found in disjunct marshes and fens in the Alps, Scandinavia and British Isles. Wolfe and colleagues (2023) reported physiological differences in photosynthesis and nutrient transport consistent with the contrasting ecological conditions preferred by the two allotetraploids. These appear to have helped the two allotetraploids adapt to different soil chemistry conditions, with *D. traunsteineri* restricted to wet, nutrient-poor soils, in particular with extremely low levels of nitrate (Wolfe et al. 2023). Furthermore, reciprocal transplant experiments have shown that much of the gene expression divergence between the sibling allopolyploids is enviromentally

144 plastic, and appeared to be associated to distinct, environmentally-driven, biotic communities
145 in their roots (Emelianova et al. 2025). In stark contrast, analysis of the genomes of several
146 *Dactylorhiza* allopolyploids found relatively little difference between *D. majalis* and *D.*
147 *traunsteineri* both in genome size and TE activity suggesting a strong control of TEs
148 following recurrent allopolyploidizations (Eriksson et al. 2022).

149 In this study, we comparatively investigate the extent of smRNAs regulation in *D. majalis* and
150 *D. traunsteineri*. Specifically, we study: (1) how does smRNA abundance in the
151 allopolyploids reflects parental expression—whether it follows additive, dominant, or
152 transgressive patterns—and how consistent are these patterns across the sibling
153 allopolyploids; (2) the degree and similarity of smRNA targeting of TEs in the two sibling
154 allotetraploids; (3) potential correlations between smRNA-mediated gene regulation and
155 distinct eco-physiological traits of *D. majalis* and *D. traunsteineri*.

156

157

158 **Materials and Methods**

159 **Sampling**

160 Adult specimens of each species of interest were transplanted from multiple localities across
161 Europe to a common garden in Vienna, Austria. *Dactylorhiza* includes perennial plants that
162 produce a new tuber yearly, which supports growth in the following year. Two years after
163 transplantation, in the morning of 14 May 2014, leaf tissue at a similar developmental stage
164 for all accessions was fixed in RNAlater (Sigma) and stored at -80°C until further processing.
165 Our final sampling included four accessions of each diploid (*D. incarnata* and *D. fuchsii*),
166 together with nine (*D. majalis*) and twelve (*D. traunsteineri*) individuals of each allotetraploid
167 (Supplementary Table S1). With few exceptions, each accession of each species originated
168 from different localities across Europe. The diploid smRNA data has been previously
169 investigated by Balao et al. (2017). Independent RNA isolations from the same tissue
170 fixations of *D. majalis* and *D. traunsteineri* were used for RNA-seq analyses by Wolfe et al.
171 (2023).

172

173 **Library preparations and sequencing**

174 Small RNA extraction, library preparation and sequencing for this experiment was previously
175 described for diploid investigations (Balao et al. 2017). In short, isolation of smRNA was
176 performed with the mirVana miRNA Isolation Kit (Life Technologies). The smRNA-enriched
177 RNA extracts were further purified by gel size selection on 15% TBE-Urea pre-casted gels

178 (Life Technologies). The small RNA libraries were prepared with the NEBNext Multiplex
179 Small RNA Library Prep Set for Illumina (New England Biolabs). Illumina sequencing as 50
180 bp single-end reads was performed at the Vienna BioCenter Core Facilities (VBCF;
181 https://www.viennabiocenter.org/).

182

183 **Read processing and quantifying**

184 The libraries were demultiplexed using BamIndexDecoder v.1.03 (available from http://wtsi-
185 npg.github.io/illumina2bam/#BamIndexDecoder). The raw reads were quality filtered,
186 trimmed from adapters used by the NEB kit, and size selected for the range 20-24 nt using
187 the CLC Genomics Workbench v.8.0 (QIAGEN). The filtered reads were mapped to the
188 available reference genome of *D. incarnata* v.1.0 (Wolfe et al. 2023) using STAR v.2.7.3a
189 (Dobin et al. 2013), with the settings --alignIntronMax 1 and --outFilterMismatchNoverLmax
190 0.05. A single pass alignment approach was used, retaining further only multi-mapping reads
191 under a threshold of 100 (--outFilterMulitmapNmax 100). A structural annotation-free
192 approach was used for the initial analysis, in order to identify any regions of the genome
193 (including non-genic) being differentially targeted (DT) by smRNA. After sorting the bam files
194 by reference coordinates with Samtools v.1.10 (Danacek et al. 2011), read coverage was
195 assessed for each individual along 100 bp windows of the genome using bamCoverage from
196 deepTools2 v.3.5.1 (Ramirez et al. 2016). All windows with a minimum coverage of ten in at
197 least one sample were retained in a bed file; neighbouring windows were merged into
198 continuous regions using bedtools merge v.2.29.2 (Quinlan and Hall 2010). The bed file
199 containing retained regions (i.e., peaks of smRNA targeting) across all individuals was then
200 converted into a gff3 file with a custom python script. Lastly, a table of counts was obtained
201 running featureCounts from the Subread v.2.0.0 package (Liao et al. 2014), accounting for
202 multiple mappers fractionally (-M --fraction), taking into account the possibility of smRNA
203 targeting in *trans*. The pipeline, scripts and raw count file, together with the peak gff3 file are
204 available at https://github.com/mc-er/dact-sRNAs.

205

206 **Differential smRNA targeting analysis**

207 Differently targeted regions were identified using R v.4.1.1 and the Bioconductor package
208 EdgeR v.3.34.0 (Robinson et al. 2010; McCarthy et al. 2012). The count matrix was filtered
209 to retain genomic regions with a minimum smRNA mapping of one count per million (CPM)
210 for a minimum of 25% of the number of samples in a group and a minimum of two samples.
211 To account for differences in library sizes the data was normalised using the weighted
212 trimmed mean of M-values (TMM) method. A quasi-likelihood negative binomial model was

implemented to find differentially targeted regions across the genome comparing either the two sibling allotetraploids or an allotetraploid towards either parental diploid. P-values were corrected for multiple testing using the Benjamini–Hochberg (BH) procedure and genomic regions only passing an FDR value of 0.05 were kept for further reporting.

Analyses targeting genomic regions around TEs were performed for reads with lengths 20–23 nt, and separately for 24 nt reads (i.e., potentially heterochromatic smRNAs, hetsiRNAs, Borges and Martienssen 2015). For analyses of genic regions all read lengths were grouped together (20–24 nt) in one analysis per comparison.

Assessing genomic interactions

Differentially targeted peaks passing a multiple testing-corrected significance threshold of 0.05 were classified as three categories of genomic interactions (Supplementary Table S2): (1) additivity—the polyploid was down regulated towards one diploid and up regulated against the other, (2) transgressive—the polyploid was up/down regulated towards both diploids, and (3) dominance—the polyploid showed no significant difference towards one diploid (the dominant) and was up/down regulated compared to the other diploid. We tested deviation from the stochastically expected number of shared peaks between the two tetraploids in each category using a hypergeometric test implemented in the Bioconductor package SuperExactTest (Wang et al. 2015).

Functional interpretation of DT genes

To get insights into a putative role for smRNA targeting in defining the divergent ecological properties between the sibling allopolyploids a gene ontology (GO) enrichment analysis of DT genes was conducted. Since the molecular effects of smRNA targeting are expected to be different depending on the respective type of targeted region, we analysed genic regions targeted by smRNAs around promoters separately from those targeted in exons or introns. Genes targeted around promoters would be expected to be down regulated or silenced, whereas those targeted by smRNAs in the gene body could experience a post-transcriptional effect or alternative splicing. Enrichment analyses for genic regions potentially affected by DT peaks were performed in R v.4.1.1, using the Bioconductor package topGO v.2.44.0 (Alexa and Rahnenfuhre 2021). We used Fisher's exact test statistics with the "weight01" algorithm to test for over-representation of specific functions. An FDR correction of p-values to account for multiple testing was applied and significance was set to a threshold of 0.05.

247 **Results**

248 **smRNAs and their genomic targets**

249 Sequencing produced an average 13.5 M reads per accession (±5.9 M SD). After quality
250 filtering and size selection, we retained an average of 5.8 M reads (±3.2 M SD) of 20–24 nt
251 (Supplementary Table S3). On average, 83.1% (±4.2% SD) of the 20–24 nt reads were
252 successfully mapped to the *D. incarnata* reference genome, with no observed mapping bias
253 across the different species (Supplementary Table S3). The distribution of read lengths
254 matched the expected profile for smRNAs (Borges and Martienssen 2015), showing modes
255 at 21–22 nt and 24 nt across all species (Fig. 2a).

256 In general, *D. fuchsii* exhibited higher CPM amounts of 20–22 nt smRNAs compared to *D.*
257 *incarnata*, although only the 21 nt smRNAs were significantly different between the two
258 diploids (Fig 2a–c). The reverse trend was observed for the 23–24 nt smRNAs, with *D.*
259 *incarnata* showing significantly higher CPM values for both 23 nt and 24 nt smRNAs
260 compared to *D. fuchsii* (Fig. 2a). The two allotetraploids were much more similar to each
261 other than the diploid species (Fig. 2a). For the 21 nt smRNAs, both allotetraploids had a
262 significantly higher CPM than their paternal species (*D. incarnata*), but significantly lower
263 CPM than the maternal species (*D. fuchsii*) (Fig. 2a). A large proportion of the 20–22 nt
264 smRNAs mapped to exonic regions, while 24 nt smRNAs mapped predominantly to TEs and
265 intergenic regions (Fig. 2b–c, Supplementary Fig. S1).

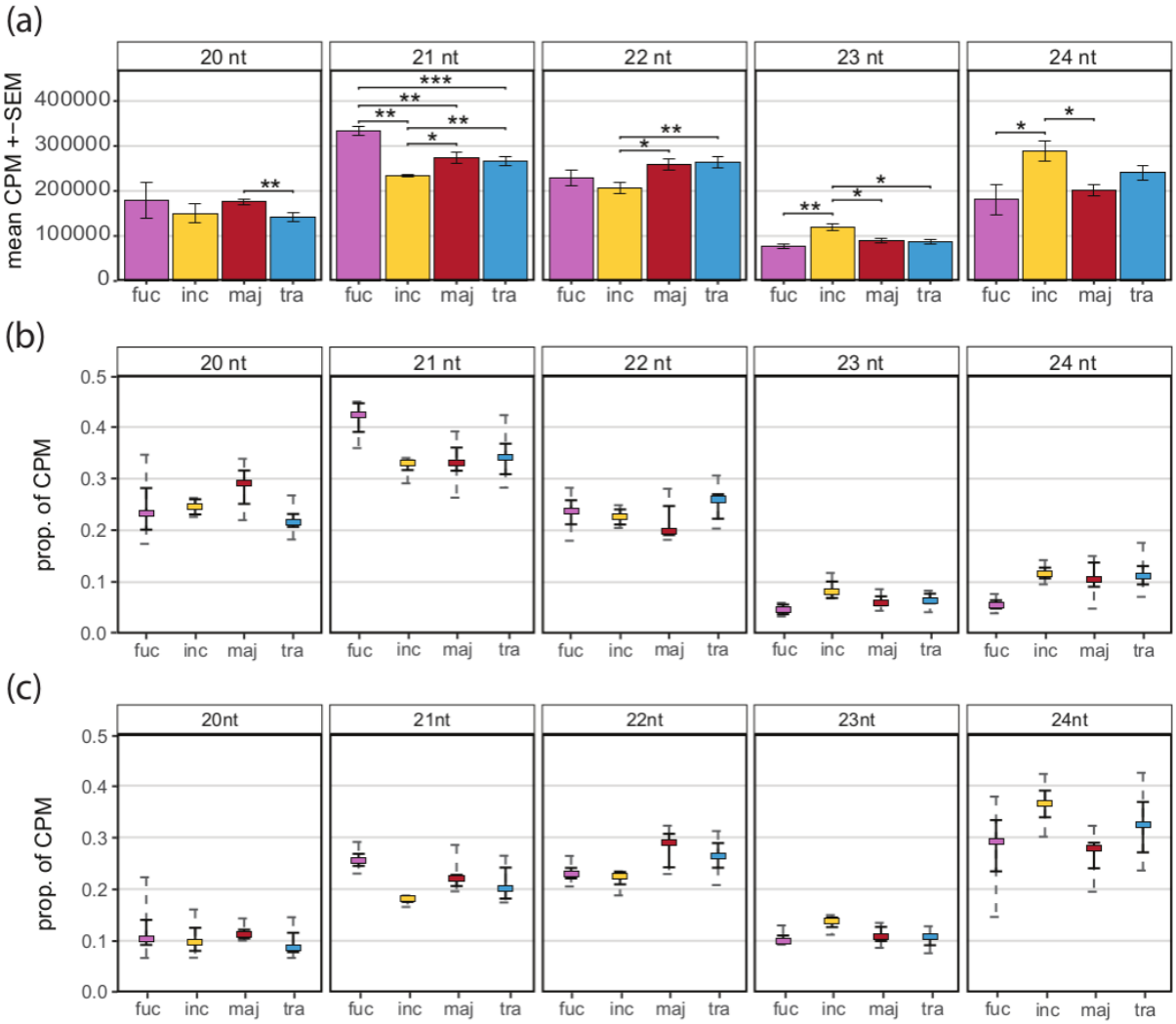
266 Our annotation-free approach identified 2,454,439 putative smRNA targeting peaks across
267 the reference genome. After filtering to retain peaks with CPM > 1 in at least two samples
268 and 25% of samples within a group, between 102,222 and 370,440 peaks were retained
269 depending on the comparison. A multidimensional scaling (MDS) plot based on all 20–24 nt
270 smRNAs reads across 212,434 peaks clearly separated the two diploid species (*D. fuchsii*
271 and *D. incarnata*), as well as the diploids from the allotetraploids, while the allotetraploids
272 overlapped centrally (Fig. 1b). Within *D. traunsteineri* a geographical clustering was apparent
273 (not shown).

274

275 **smRNA activity in ecologically-divergent allopolyploids**

276 We first compared the two ecologically divergent sibling allotetraploids using the combined
277 set of 20–24 nt reads, retaining 152,503 peaks that passed the CPM filtering threshold
278 (Supplementary Table S4). Among these, we identified 2,106 DT peaks between *D.*
279 *traunsteineri* and *D. majalis*, of which approximately half (1,016 peaks) were associated with
280 genes. Gene-associated peaks were defined as those overlapping a gene model (including

281 exons and introns) or within 1,000 bp upstream of a gene model (hereafter referred to as
282 'promoters') (Supplementary Table S4).



283

284 **Figure 2.** smRNA reads for each species divided into length groups. **(a)** shows the mean
285 count per million (CPM) read count with error bars representing standard error of the mean.
286 Horizontal lines indicate the significant differences between species with stars representing
287 thresholds of * - 0.05, ** - 0.01, *** - 0.001. **(b)** and **(c)** show the proportion of reads for each
288 species and read length based mapping to genomic features, i.e., (b) to exons and (c) to
289 TEs.

290

291 The majority of genic DT smRNA peaks (i.e., 762) were located within introns. A smaller
292 number (82) overlapped multiple genic features (intron, exon, and/or promoters)
293 (Supplementary Table S4). Across all genic categories, at least three times more smRNA
294 peaks were over-targeted in the older allotetraploid *D. majalis* compared to *D. traunsteineri*
295 (Supplementary Table S4, Fig. 3a–c).

296 Comparison of DT peaks with previously identified differentially expressed (DE) genes from
297 the same experiment (Wolfe et al. 2023) revealed that most genes DT by smRNAs did not

show differential expression between the two allopolyploids. This suggests that smRNAs may be involved in post-transcritional regulation or alternative splicing, rather than transcriptional silencing or activation (Fig. 3f, Supplementary Table S5).

GO enrichment analysis of genes DT by smRNAs between the ecologically divergent allotetraploids revealed an asymmetric overrepresentation of functional categories, with more GO terms enriched among genes more highly targeted in *D. majalis* than in *D. traunsteineri* (Fig. 5a, Supplementary Tables S6–S7). The vast majority of enriched terms were found among genes targeted within gene bodies, and were associated with functions related to regulation of transcription and translation, transmembrane transport, response to biotic and abiotic stress, DNA repair, and regulation of meiosis and/or mitosis (Supplementary Table S7).

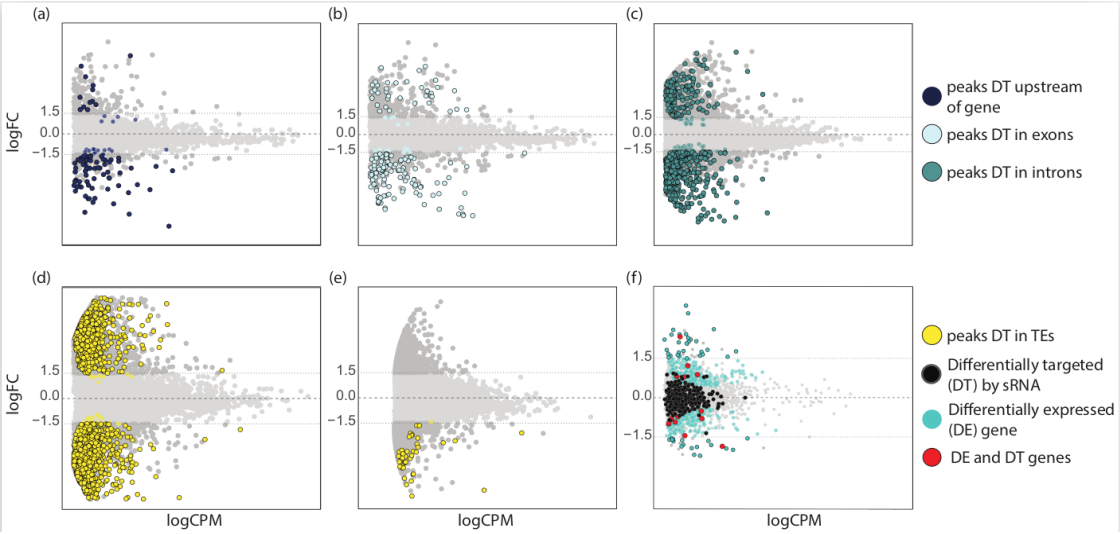


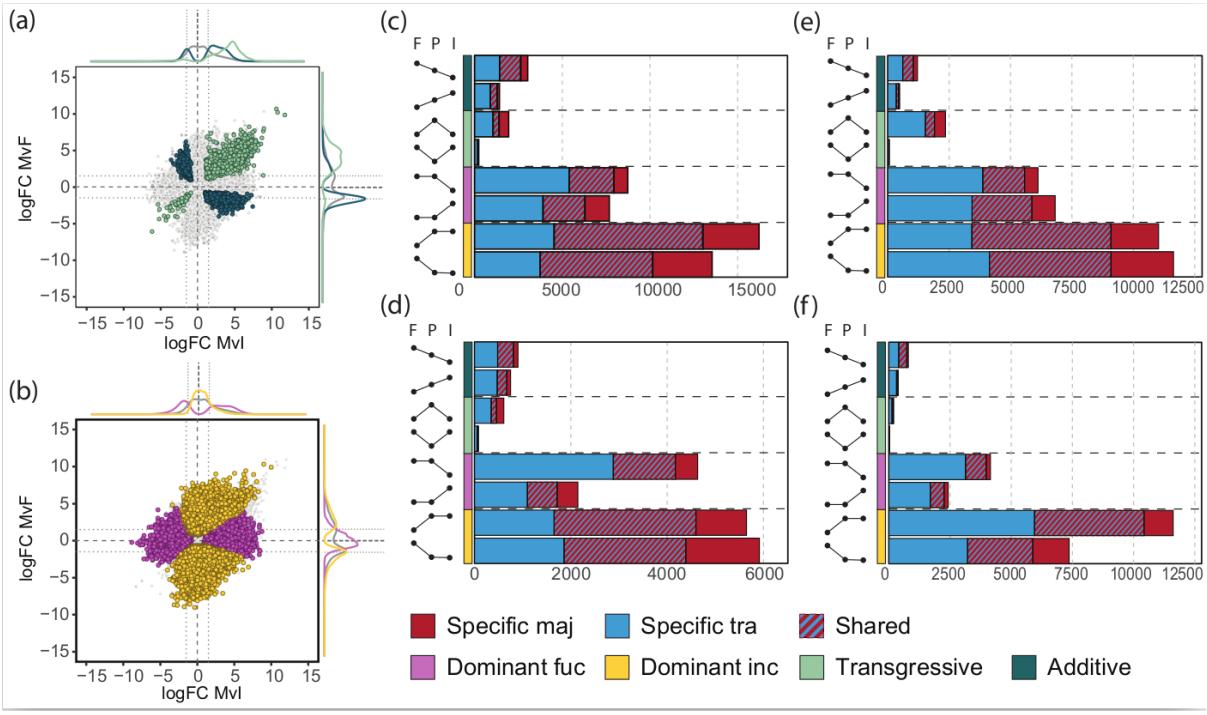
Figure 3. MA plots showing differences between *D. majalis* and *D. traunsteineri*, the central dashed line marks 0 while the dotted lines mark logFC -1.5 and 1.5. Negative logFC values represent over-targeting in *D. majalis* as compared to *D. traunsteineri*, whereas positive logFC values indicate over-targeting in *D. traunsteineri*. **(a–c)** shows peaks differently targeted for genes targeted by 20–24 nt smRNAs in (a) promoter/1,000 bp upstream gene models, (b) exons and (c) introns. **(d)** shows TEs targeted by 20–23 nt smRNAs and **(e)** TEs targeted by 24 nt smRNAs. **(f)** An MA plot of RNA-seq data (Wolfe et al. 2023), indicating according to the legend genes differentially expressed (DE) and/or DT by smRNAs.

smRNA targeting and genomic interactions

We further compared the 20-24 nt smRNA targeting patterns of each sibling allotetraploid to those of the parental diploid species. After filtering, 212,434 peaks were retained for this analysis (Supplementary Table S4). Overall, both allotetraploids showed similar targeting patterns when compared to the diploids, although the younger *D. traunsteineri* displayed a more divergent smRNA landscape than *D. majalis* (Supplementary Table S4, Supplementary Fig. S2). In both cases, the allotetraploids also showed greater divergence from the maternal

326 parent (*D. fuchsii*) than from the paternal *D. incarnata* (Supplementary Table S4,
327 Supplementary Fig. S2).

328 We characterised DT smRNAs peaks into four interaction types: additive, transgressive, and
329 dominant towards either diploid parent. The majority of peaks showed no change between
330 the allotetraploids and either diploid. Among those with DT, non-additive patterns were most
331 frequent, with dominance being the most common, particularly towards *D. incarnata*
332 (Supplementary Table S8, Fig. 4a–f, Supplementary Fig. S3–S4). Notably, in intergenic
333 regions lacking annotated TEs, we observed a higher proportion of smRNA peaks showing
334 dominance towards *D. fuchsii* than observed for any other genomic contexts (Supplementary
335 Table S8, Fig 4d, Supplementary Fig. S3). Among the transgressive peaks, both
336 allotetraploids most frequently showed over-targeting by smRNAs in comparison to both
337 diploid parents (Supplementary Table S8, Fig. 4a, Fig. 4c–f, Supplementary Fig. S3–S4). A
338 notable general pattern over all types of genomic interactions was that a large proportion of
339 DT smRNA peaks were shared between the two sibling allotetraploids (Supplementary Table
340 S8, Fig. 4b–c, Supplementary Fig. S3–S4), despite their independent evolutionary histories
341 (Brandrud et al. 2020; Hawranek 2021).



342

343 **Figure 4. (a–b)** Patterns of genomic interactions comparing logFC values for *D. majalis*
344 towards either diploid, *D. fuchsii* on the Y-axis and *D. incarnata* on the X-axis for all peaks
345 (i.e., 20–24 nt smRNAs) in annotated genes. **(a)** shows with colored symbols transgressive
346 and additive patterns; whereas **(b)** dominant to either diploid, according to the legend. **(c-f)**
347 Number of DT peaks in each type of genomic interaction with peaks in *D. traunsteineri* in
348 blue and *D. majalis* in red; the striped part indicates the portion of DT peaks shared between
349 the allotetraploids. The two annotation columns along the Y-axis represent the type
350 (coloured bars according to legend) of genomic interaction while dots connected with lines

351 show the direction. Letters above each column of dots represent I - *D. incarnata,* P -
352 polyploid and F - *D. fuchsii.* **(c)** 20–24 nt smRNAs associated to genic regions, **(d)** 20–24 nt
353 smRNA peaks outside annotated genic and TE regions, **(e)** 20–23 nt smRNAs in TEs and **(f)**
354 24 nt smRNAs in TEs.

355

356 We also examined the distribution of interaction categories across genic regions (promoters,
357 exons, introns). In *D. majalis,* nerly 75% promoter-associated DT peaks showed dominance
358 towards *D. incarnata*, and almost 50% showed lower smRNA targeting than *D. fuchsii*
359 (Supplementary Fig. S5). In contrast, *D. traunsteineri* exhibited greater dominance toward *D.*
360 *fuchsii* across all genic regions compared to *D. majalis* (Supplementary Fig. S5).

361 For genes transgressively targeted by smRNAs, similar functional categories were enriched
362 in both allotetraploids, including response to stress, transport, regulation of
363 transcription/translation, and regulation of meiosis/mitosis. However, these enrichments
364 involved largely different genes and pathways in each species (Fig. 5b–c, Supplementary
365 Tables S9–S12). In *D. traunsteineri*, transgressive smRNA targeting was associated with
366 enrichment for cell cycle regulation and stress response functions, and the respective genes
367 were targeted within exons or introns (Fig. 5b, Supplementary Tables S11–S12). In *D.*
368 *majalis*, transgressive peaks occurred mostly within introns, and enrichments were
369 associated with regulation of gene function (Fig. 5c, Supplementary Tables S9–S10).

370

371 **smRNAs targeting TEs**

372 For the TE-focused analyses, we divided the data into two sets: (a) 24 nt smRNAs, analysed
373 separately since they should be predominantly het-siRNAs and therefore target more
374 frequently TEs (Fig2c, Supplementary Fig. S1); and (b) 20-23 nt smRNAs, grouped together.
375 After CPM filtering, for the comparative analysis of *D. majalis* versus *D. traunsteineri* more
376 peaks passed the threashold for 24 nt smRNAs (315,838) than for 20–23 nt smRNAs
377 (102,222) (Supplementary Table S4). However, the DT analysis using a generalised linear
378 model uncovered the opposite trend: 3,242 DT peaks were identified for the 20–23 nt
379 smRNAs, compared to only 103 DT peaks for the 24 nt smRNAs (Supplementary Table S4,
380 Fig. 3d–e) suggesting a consistent targeting of TEs with het-siRNAs among the
381 allopolyploids. Very few TE-associated peaks were DT between the allotetraploids in both
382 smRNA size classes simultaneously (Supplementary Fig. S6). In both datasets,
383 approximately half of the DT peaks overlapped annotated TEs (Supplementary Table S4).
384 Many of these were also located within gene introns, while a smaller number were found in
385 the vicinity of gene promoters (Supplementary Table S4).

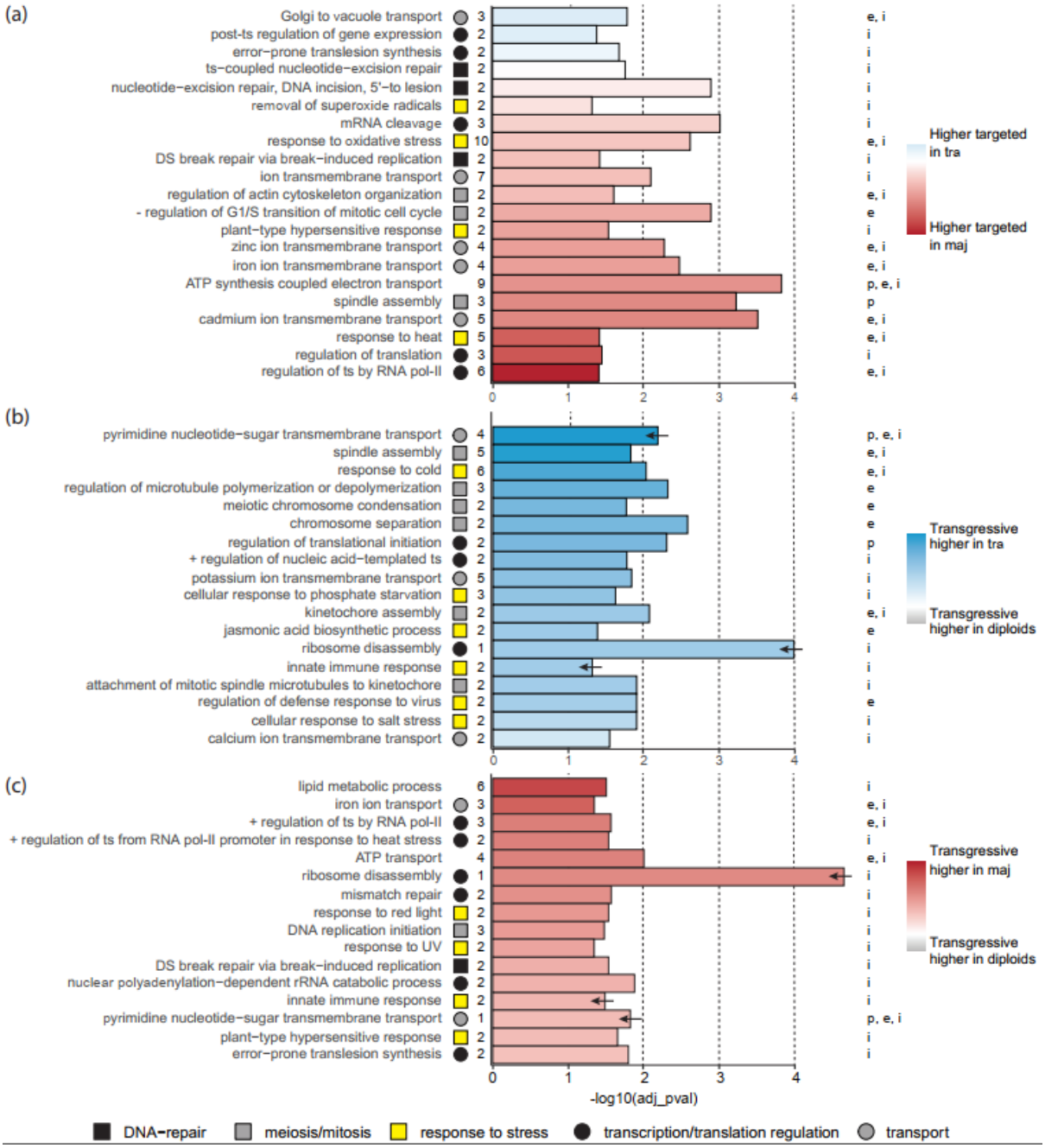


Figure 5. Summary of representative functional enrichments for **(a)** the allopolyploid comparison, **(b)** transgressive smRNA targeting of genic regions in *D. traunsteineri*, and **(c)** transgressive targeting in *D. majalis*. The size of the bars shows the adjusted significance of the enriched GO term. For the full result see Supplementary Tables S7, S10 and S12. The color of the bars indicates the z-score, computed based on the logFC of the respective targeting, according to the scale indicated in the legends. Arrows on bars in the transgressive GO enrichments (panels b and c) indicate terms present in both comparisons. Symbols next to the GO terms indicate the broad category to which the GO terms pertain, according to the legend. Numbers next to symbols indicate the number of enriched genes. Letters at the right of each plot indicate the genetic region targeted by smRNAs; p - 'promotor', e - exon, i - intron. Abbreviations used in GO terms: + positive, - negative, DS double strand, pol polymerase, ts transcription.

400 Similar to the results for the genic regions, *D. majalis* exhibited more frequent over-targeting
401 of TEs than *D. traunsteineri* among DT peaks, with the exception of 20–23 nt peaks
402 associated with TEs located around promoters, where *D. traunsteineri* showed increased
403 targeting (Supplementary Table S4). The majority of TE-associated DT peaks were located
404 within LTR retrotransposons, particularly those classified as Ty3-*gypsy* (Supplementary Fig.
405 S7–S8). Among smRNAs targeting Ty3-*gypsy* elements, a slightly higher number targeted
406 non-chromoviruses than chromoviruses; however, most such elements were annotated only
407 generally as Ty3-*gypsy* in the *D. incarnata* reference genome (Supplementary Fig. S8).

408 We also compared smRNA targeting of TEs in the allotetraploid genomes against their
409 parental diploids, again investigating 20–23 nt and 24 nt smRNAs separately. After CPM
410 filtering, we retained 139,736 and 370,440 peaks for 20–23 nt and 24 nt smRNAs,
411 respectively (Supplementary Table S4). As with genic targets, we classified the DT peaks
412 into four categories of genomic interaction: additive, transgressive, and dominant towards
413 either diploid. In both datasets, a significant number of shared DT peaks were found
414 between *D. majalis* and *D. traunsteineri* across interaction categories ($p < 0.001$,
415 hypergeometric test; Fig 3e-f, Supplementary Fig. S3, Fig. 4e–f, Supplementary Fig. S9–
416 S10).

417 Patterns of dominance were again prevalent, with more peaks dominant towards either
418 diploid than additive or transgressive. For 20–23 nt smRNAs, slightly more peaks showed
419 dominance toward the diploid with lower smRNA targeting, whereas for 24 nt smRNAs the
420 opposite trend was observed (Fig. 3e–f, Supplementary Fig. S3). The contrast between
421 dominant types (higher vs. lower targeting) was less pronounced for 20–23 nt and more
422 distinct for 24 nt smRNAs (Fig. 3e–f, Supplementary Fig. S3). Three patterns observed for
423 gene targeting were also evident in TE targeting: First, for most additive peaks, the smaller
424 genome of *D. fuchsii* exhibited higher smRNA targeting than the larger *D. incarnata* (Fig. 3e–
425 f, Supplementary Fig. S3). Second, for transgressive peaks, both allotetraploids generally
426 showed elevated smRNA targeting compared to the diploids (Fig. 3e–f, Supplementary Fig.
427 S3, Supplementary Fig. S9–S10). Last, *D. traunsteineri* displayed greater differences to the
428 parental diploids than D. majalis (Fig. 3e–f, Supplementary Fig. S3, Supplementary Fig. S9–
429 S10).

430

431 **Discussion**

432 In this study, we provide a comprehensive analysis of the contribution of smRNAs to
433 genomic and transcriptomic diversification following repeated allopolyploidization in
434 *Dactylorhiza majalis* s.l. We examined smRNA expression in two established sibling

435 allotetraploids and their diploid progenitors, all grown under common garden conditions,
436 thereby minimizing environmental variation. Our primary objectives were to comparatively
437 characterize (*i*) patterns of genomic interaction in the allopolyploids and (*ii*) smRNA-
438 mediated regulation of TEs and gene activity.

439

440 **Consistent dominance in the allopolyploids towards the paternal and larger genome**

441 Allopolyploidization merges divergent genomes, giving rise to novel genetic and regulatory
442 interactions. Common outcomes include epistasis and non-additive genetic or epigenetic
443 effects, such as smRNA-mediated gene regulation and TE silencing. Our results reveal a
444 consistent dominance of smRNA targeting aligned to the paternal, larger subgenome (*D.*
445 *incarnata*), especially for 24 nt TE-associated smRNAs. This pattern was observed across
446 both sibling allotetraploids (Fig. 4, Supplementary Fig. S5), resulting in a comparatively small
447 number of DT peaks between them. Similar non-additive smRNA responses following
448 hybridization and/or whole-genome duplication have been documented in *Arabidopsis* (Ha et
449 al. 2009), *Brassica* (Fu et al. 2016; Martinez Palacios et al. 2019), and *Spartina* (Cavé-Radet
450 et al. 2020), though it has remained unclear before whether such patterns remain consistent
451 across evolutionary timescales.

452 Given that *D. incarnata* has a larger genome and higher TE load than *D. fuchsii* (Eriksson et
453 al. 2022), our findings align well with previous reports that smRNA-mediated silencing often
454 preferentially targets the TE-rich subgenome, as observed in wheat (Woodhouse et al. 2014)
455 and *Brassica rapa* (Cheng et al. 2016). In stark contrast, gene expression dominance in
456 allopolyploids frequently shifts toward the subgenome with fewer TEs, a pattern that can
457 arise rapidly after allopolyploidization (for example in *Mimulus*; Edger et al. 2017) and persist
458 across multiple generations (Gaebelein et al. 2019; Liu and Wang 2023).

459 While our data indicate a stronger dominance towards *D. incarnata*, we acknowledge that
460 some of these patterns may be influenced by reference genome bias, as we only had access
461 to a *D. incarnata* reference genome, limiting our ability to assess homeolog-specific
462 targeting. Due to the short lengths of smRNAs and their targeting with mismatches in plants,
463 even if references for both genomes would be available, studying subgenome effects at the
464 level of smRNAs targeting remains a challenge. Nonetheless, in this study mapping success
465 was comparable across samples (ca. 83%; Supplementary Table S3), and intergenic
466 smRNA dominance towards *D. fuchsii* was also observed, suggesting that reference bias
467 alone cannot account for our observations. Still, lineage-specific degradation or divergence
468 of TEs in *D. fuchsii* may have reduced smRNA detection from this genome. Notably, *D.*
469 *incarnata* and *D. fuchsii* diverged approximately 5.5 million years ago (Brandrud et al. 2020).

470

471 **Divergent regulation of TEs by 24 nt and 20-23 nt smRNAs**

472 As expected, 24 nt smRNAs—typically het-siRNAs—were strongly associated with TEs,
473 particularly LTR retrotransposons (Fig. 2c, Supplementary Fig. S1). However, the number of
474 DT peaks for 24 nt smRNAs between the allotetraploids was surprisingly low, suggesting
475 that TE silencing proceeds in a relatively consistent manner across independently evolved
476 lineages. In contrast, 20–23 nt smRNAs exhibited markedly greater divergence between the
477 sibling allopolyploids (Fig. 3d–e), consistent with a more dynamic, lineage-specific regulatory
478 role.

479 Previous studies have shown that TE-associated smRNA activity often peaks shortly after
480 hybridization or whole-genome duplication, followed by stabilization through methylation or
481 other epigenetic mechanisms over the next generations (Jiao et al. 2018; Martinez Palacios
482 et al. 2019). In *Arabidopsis,* Ha et al. (2009) suggested that allopolyploids take advantage of
483 smRNAs inherited from the parents to buffer initial genomic shock and promote rapid
484 genome stabilization. The relatively lower divergence in 24 nt smRNA targeting observed in
485 our study may reflect the advanced age of these natural allotetraploids—approximately 74k
486 (*D. traunsteineri*) and 104k generations (*D. majalis*) (Hawranek 2021). Nonetheless, for the
487 few DT peaks detected, smRNAs abundance was higher in the older *D. majalis* (Fig. 3),

488 possibly reflecting continued regulatory refinement associated with its broader ecological
489 range or more advanced genome diploidization.

490 Intriguingly, most smRNA differences associated with TE regulation in the ecologically
491 divergent allotetraploids involved 20–23 nt smRNAs, which likely include miRNAs and nat-
492 siRNAs. These classes are frequently implicated in responses to biotic and abiotic stress (Ito
493 et al. 2011; Marí-Ordóñez et al. 2013; Creasey et al. 2014; Lee et al. 2020). Elevated
494 expression of miRNAs (typically 20–22 nt long) has also been observed in resynthesized
495 allopolyploids relative to their diploid progenitors in several systems, including *Arabidopsis*,
496 *Brassica*, and others (Ha et al. 2009; Kenan-Eichler et al. 2011; Ghani et al. 2014),
497 underscoring the potential role of shorter smRNAs in stress adaptation and dynamic gene
498 regulation in polyploids.

499 Despite *Ty1-copia* being the most abundant LTR-TE in the *Dactylorhiza* genomes (Eriksson
500 et al. 2022), we found more 20–23 nt smRNAs mapping to, and possibly being derived from,
501 *Ty3-gypsy* elements (Supplementary Figs. S6–S7), particularly to non-chromovirus
502 subgroups. Notably, *Ty3-gypsy* elements displayed contrasting patterns between the diploid
503 progenitors: chromovirus subgroups were more prevalent in *D. fuchsii*, while non-
504 chromoviruses were more common in *D. incarnata* (Eriksson et al. 2022). The elevated

505 targeting of *Ty3-gypsy* elements by shorter smRNAs in the allopolyploids may reflect
506 subgenome conflict or TE activation following hybridization. A similar phenomenon has been
507 described in *Gossypium*, where TEs originally present in the higher-TE A-subgenome
508 transposed into the D-subgenome post-allopolyploidization (Chen et al. 2020), suggesting
509 that inter-subgenomic TE movement may contribute to regulatory instability and
510 diversification.

511

512 **Genic smRNA targeting and regulatory implications**

513 Among others, polyploid formation causes meiotic/mitotic challanges and genetic
514 redundancy. On one hand, even if allopolyploids have two full sets of divergent genomes,
515 some chromosomes or chromosomal regions may still retain homology; hence it becomes
516 crucial to sort homologous and homeologous to ensure correct chromosome segregation
517 during meiosis. Preferential pairing of homologs has been observed in early stages of
518 polyploid formation in *Arabidopsis suecica* (Pecinka et al. 2011) and *Brassica napus*
519 (Grandont et al. 2014), although elevated rates of homeolog related errors have also been
520 observed in the latter system (Xiong et al. 2021). On the other hand, given their multiple sets
521 of gene copies, allopolyploids must enforce genetic diploidization by suppressing redundant
522 or misregulated genes (Lee and Chen 2001; Kashkush et al. 2002; Koh et al. 2010).
523 Duplicated copies of some genes may be retained functional if they confer adaptive benefits,
524 become specialised in different tissues or developmental stage through subfunctionalization,
525 or acquire a new function through neofunctionalization (Adams et al. 2003; Blanc and Wolfe
526 2004; Duarte et al. 2006; De Smet et al. 2013).

527 In addition to silencing TEs, smRNAs—particularly 21–22 nt species—also regulate gene
528 expression via PTGS or TGS (Borges and Martienssen 2015). We observed widespread
529 smRNA targeting of genic regions, especially exons and 1 kb upstream of gene models
530 (putative promoters), with 21 nt smRNAs contributing the most. While such targeting may
531 mediate transcript degradation, alternative splicing, or cis-silencing (Guleria et al. 2011;
532 Moss 2001), it can also represent regulatory adjustments following genome duplication, as
533 initial steps during the diploidization process.

534 Strikingly, substantial genic smRNA targeting was also detected in the diploids, especially *D.*
535 *fuchsii*. This may reflect stress-induced smRNA responses in the common garden
536 experiment, which was conducted outdoors in Vienna, Austria—a general area where only
537 *D. incarnata* occurs naturally. Environmental mismatch may have triggered stress responses
538 in *D. fuchsii*, activating miRNA- and nat-siRNA-mediated regulation. Stress-induced
539 upregulation of these types of smRNAs has been widely reported (Zhang et al. 2012; Mao et

540 al. 2021), and recent reviews (e.g., Tiwari and Rajam 2022) highlight their role in coping with
541 drought, salt, nutrient, and temperature stress across many plant systems.

542 Transgressive smRNA expression compared to diploid progenitors could offer ways in which
543 allopolyploids respond on one side to their duplicated state, and on the other side, achieve
544 novel ecological features, helping them to establish in niches outside of parental range. We
545 observed more transgressive activity at peaks targeted by 20–23 nt smRNAs—compared to
546 24 nt—especially when the allotetraploids showed stronger regulation compared to both
547 diploid parental species (Fig. 4e–f).

548 Transgressively smRNAs targeted genes were often enriched for functions related to
549 mitosis, meiosis, gene regulation, stress responses, and transport (Fig. 5). While these
550 categories were largely shared across the two allotetraploids, the individual genes and
551 pathways involved differed, suggesting divergent evolutionary routes. In the younger *D.*
552 *traunsteineri*, enrichment of smRNA targeting was observed for genes involved in cell cycle
553 and meiotic processes, possibly reflecting ongoing resolution of meiotic instability. By
554 contrast, the older *D. majalis* showed enrichments for genes involved in transcriptional and
555 translational regulation, likely reflecting a later phase of polyploid evolution focused on
556 dosage control and gene expression fine-tuning.

557 Together, these observations support a two-phase model of smRNA-mediated regulatory
558 evolution following allopolyploidization: an early phase dominated by stabilization of
559 chromosome behavior during meiosis, and a later phase involving refinement of gene
560 redundancy and functional divergence. Such dynamics have been proposed in other
561 systems including *Arabidopsis*, *Brassica*, and *Gossypium* (Ha et al. 2009; Fu et al. 2016;
562 Chen et al. 2020). In *Dactylorhiza* we observe (i) more differences relative to diploid parents
563 for *D. traunsteineri*, consistent with an earlier stage of regulatory resolution, and (ii) a higher
564 number of upregulated smRNA peaks in *D. majalis*, suggesting more established and
565 efficient regulatory networks.

566

567 **Conclusions and Future Directions**

568 This study underscores the dual role of small RNAs (smRNAs) in allopolyploid genome
569 evolution—both in the stabilization of the genome via transposable element (TE) silencing
570 and in the regulation of gene expression. Despite their shared ancestry, *D. majalis* and *D.*
571 *traunsteineri* exhibit clear lineage-specific patterns of smRNA activity, particularly in genic
572 regions, suggesting divergent regulatory trajectories following allopolyploidization.

Due to their rapid action, diversity, and regulatory plasticity, smRNAs are likely central players in the early and long-term genomic reorganization of allopolyploid species. Our results support a model in which smRNAs mediate both immediate stabilization and longer-term functional divergence after genome merging.

Future work should integrate smRNA expression profiles with homeolog-specific reference genomes and DNA methylation landscapes. Such approaches would provide deeper insights into subgenome-specific regulation, the dynamics of smRNA–target interactions, and the evolutionary consequences of smRNA-mediated gene and TE control in natural allopolyploids.

Acknowledgements

This work was financially supported by the Austrian Science Fund (FWF) through grant Y661-B16 awarded to O.P., and through the doctoral programme (DK) grant W1225-B20, awarded to a faculty team including O.P. Faculty members, in particular Andreas Futschik, Robert Kofler and Christian Schlötterer, as well as the students and the scientific advisory board of the Vienna Graduate School of Population Genetics (www.popgen-vienna.at) are gratefully acknowledged for their valuable discussions and feedback throughout this work. We thank Richard Bateman for assistance during planning and fieldwork, and Juliane Baar, Daniela Paun and Maria Teresa Lorenzo for their support with laboratory procedures. Sequencing was performed at the Vienna BioCenter CoreFacilities (VBCF; https://www.viennabiocenter.org/). Computational resources were provided by the Life Science Compute Cluster (LiSC) of the University of Vienna. We also acknowledge Natural England and the Forestry Commission (UK), and regional county administrations in Austria, France, and Sweden for issuing the necessary collecting permits.

Author contributions

Study conceived and designed by OP. Sampling and transplantation to the common garden conducted by OP, FB, MH, MCE and TMW. Read processing and mapping done by MT. Identification of putative smRNA peaks across the genome MT and MCE. Data analyses of differentially targeted genes and TEs conducted by MCE. Analysis of overlapping differentially expressed and differentially targeted genes done by TMW. Interpretation of the results was undertaken by MCE. The manuscript was drafted by MCE, with feedback from OP, and was revised and approved by all authors.

606 **References**

607 Adams KL, Cronn R, Percifield R, Wendel JF. Genes duplicated by polyploidy show unequal
608 contributions to the transcriptome and organ-specific reciprocal silencing. Proc Natl Acad Sci
609 U S A. 2003;100(8):4649–54. doi:10.1073/pnas.0630618100

610 Alexa A, Rahnenfuhrer J. topGO: Enrichment Analysis for Gene Ontology. R Package
611 Version 2.44.0. https://bioconductor.org/packages/topGO. Accessed 15 May 2025.

612 Ariel FD, Manavella PA. When junk DNA turns functional: transposon-derived non-coding
613 RNAs in plants. J Exp Bot. 2021;72(11):4132–41. doi:10.1093/jxb/erab082

614 Balao F, Trucchi E, Wolfe TM, Hao BH, Lorenzo MT, Baar J, et al. Adaptive sequence
615 evolution is driven by biotic stress in a pair of orchid species (Dactylorhiza) with distinct
616 ecological optima. Mol Ecol. 2017;26(14):3649–62. doi:10.1111/mec.14123

617 Blanc G, Wolfe KH. Functional divergence of duplicated genes formed by polyploidy during
618 Arabidopsis evolution. Plant Cell. 2004;16(7):1679–91. doi:10.1105/tpc.021410

619 Borges F, Martienssen RA. The expanding world of small RNAs in plants. Nat Rev Mol Cell
620 Biol. 2015;16(12):727–41. doi:10.1038/nrm4085

621 Brandrud MK, Baar J, Lorenzo MT, Athanasiadis A, Bateman RM, Chase MW, et al.
622 Phylogenomic relationships of diploids and the origins of allotetraploids in Dactylorhiza. Syst
623 Biol. 2020;69(1):91–109. doi:10.1093/sysbio/syz035

624 Carthew RW, Sontheimer EJ. Origins and mechanisms of miRNAs and siRNAs. Cell.
625 2009;136(4):642–55. doi:10.1016/j.cell.2009.01.035

626 Cavé-Radet A, Giraud D, Lima O, El Amrani A, Aïnouche ML, Salmon A. Evolution of small
627 RNA expression following hybridization and allopolyploidization: insights from Spartina
628 species (Poaceae, Chloridoideae). Plant Mol Biol. 2020;102:55–72. doi:10.1007/s11103-
629 019-00931-w

630 Cavé-Radet A, Salmon A, Tran Van Canh L, et al. Recent allopolyploidy alters Spartina
631 microRNA expression in response to xenobiotic-induced stress. Plant Mol Biol.
632 2023;111:309–28. doi:10.1007/s11103-022-01328-y

633 Chen X, Rechavi O. Plant and animal small RNA communications between cells and
634 organisms. Nat Rev Mol Cell Biol. 2021; doi:10.1038/s41580-021-00425-y

635 Chen ZJ, Sreedasyam A, Ando A, Song Q, De Santiago LM, Hulse-Kemp AM, et al.
636 Genomic diversifications of five Gossypium allopolyploid species and their impact on cotton
637 improvement. Nat Genet. 2020;52:525–33. doi:10.1038/s41588-020-0614-5

638 Cheng F, Sun C, Wu J, Schnable J, Woodhouse MR, Liang J, et al. Epigenetic regulation of
639 subgenome dominance following whole genome triplication in Brassica rapa. New Phytol.
640 2016;211(1):288–99. doi:10.1111/nph.13884

641 Creasey KM, Zhai J, Borges F, Van Ex F, Regulski M, Meyers BC, Martienssen RA. miRNAs
642 trigger widespread epigenetically activated siRNAs from transposons in Arabidopsis. Nature.
643 2014;508:411–15. doi:10.1038/nature13069

644 Danecek P, Auton A, Abecasis G, Albers CA, Banks E, Depristo MA, et al. The variant call
645 format and VCFtools. Bioinformatics. 2011;27(15):2156–58.
646 doi:10.1093/bioinformatics/btr330

647 De Smet R, Adams KL, Vandepoele K, Van Montagu MCE, Maere S, Van de Peer Y.
648 Convergent gene loss following gene and genome duplications creates single-copy families
649 in flowering plants. Proc Natl Acad Sci U S A. 2013;110(8):2898–903.
650 doi:10.1073/pnas.1300127110

651 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast
652 universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21.
653 doi:10.1093/bioinformatics/bts635

654 Duarte JM, Cui L, Wall PK, Zhang Q, Zhang X, Leebens-Mack J, et al. Expression pattern
655 shifts following duplication indicative of subfunctionalization and neofunctionalization in
656 regulatory genes of Arabidopsis. Mol Biol Evol. 2006;23(2):469–78.
657 doi:10.1093/molbev/msj051

658 Edger PP, Smith R, McKain MR, Cooley AM, Vallejo-Marín M, Yuan Y, et al. Subgenome
659 dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally
660 established neo-allopolyploid monkeyflower. Plant Cell. 2017;29(9):2150–67.
661 doi:10.1105/tpc.17.00010

662 Emelianova K, Hawranek A-S, Eriksson M, Wolfe TM, Paun O. Ecological divergence of
663 sibling allopolyploid marsh orchids is associated with species specific plasticity and distinct
664 fungal communities. Plant J. 2025;121:e70001.

665 Eriksson MC, Mandáková T, McCann J, Temsch EM, Chase MW, Hedrén M, et al. Repeat
666 dynamics across timescales: a perspective from sibling allotetraploid marsh orchids
667 (Dactylorhiza majalis s.l.). Mol Biol Evol. 2022;39(8):msac167. doi:10.1093/molbev/msac167

668 Fedoroff NV. Transposable elements, epigenetics, and genome evolution. Science.
669 2012;338:758–67. doi:10.1126/science.338.6108.758

670 Fu Y, Xiao M, Yu H, Mason AS, Yin J, Li J, et al. Small RNA changes in synthetic Brassica
671 napus. Planta. 2016;244(3):607–22. doi:10.1007/s00425-016-2529-z

672 Fultz D, Choudury SG, Slotkin RK. Silencing of active transposable elements in plants. Curr
673 Opin Plant Biol. 2015;27:67–76. doi:10.1016/j.pbi.2015.05.027

674 Gaebelein R, Schiessl SV, Samans B, Batley J, Mason AS. Inherited allelic variants and
675 novel karyotype changes influence fertility and genome stability in Brassica allohexaploids.
676 New Phytol. 2019; doi:10.1111/nph.15804

677 Ghani MA, Li J, Rao L, Raza MA, Cao L, Yu N, et al. The role of small RNAs in wide
678 hybridisation and allopolyploidisation between Brassica rapa and Brassica nigra. BMC Plant
679 Biol. 2014;14(1):272. doi:10.1186/s12870-014-0272-9

680 Grandont L, Cuñado N, Coriton O, Huteau V, Eber F, Chèvre AM, et al. Homoeologous
681 chromosome sorting and progression of meiotic recombination in Brassica napus: Ploidy
682 does matter! Plant Cell. 2014;26:1448–63. doi:10.1105/tpc.114.122788

683 Guleria P, Mahajan M, Bhardwaj J, Yadav S. Plant small RNAs: biogenesis, mode of action
684 and their roles in abiotic stresses. Genomics Proteomics Bioinformatics. 2011;9:183–99.
685 doi:10.1016/S1672-0229(11)60022-9

686 Ha M, Lu J, Tian L, Ramachandran V, Kasschau KD, Chapman EJ, et al. Small RNAs serve
687 as a genetic buffer against genomic shock in Arabidopsis interspecific hybrids and
688 allopolyploids. Proc Natl Acad Sci U S A. 2009;106(42):17835–40.
689 doi:10.1073/pnas.0907003106

690 Hassan AH, Mokhtar MM, El Allali A. Transposable elements: multifunctional players in the
691 plant genome. Front Plant Sci. 2024;14:1330127. doi:10.3389/fpls.2023.1330127

692 Hawranek A-S. The impact of recurrent origins and gene flow on the genetic structure of
693 allopolyploid marsh orchids (Dactylorhiza, Orchidaceae). MSc thesis. University of Vienna;
694 2021. doi:10.25365/thesis.70360

695 Ito H, Gaubert H, Bucher E, Mirouze M, Vaillant I, Paszkowski J. An siRNA pathway
696 prevents transgenerational retrotransposition in plants subjected to stress. Nature.
697 2011;472(7341):115–20. doi:10.1038/nature09861

698 Jiang X, Song Q, Ye W, Chen ZJ. Concerted genomic and epigenomic changes accompany
699 stabilization of Arabidopsis allopolyploids. Nat Ecol Evol. 2021;5(10):1382–93.
700 doi:10.1038/s41559-021-01523-y

701 Jiao W, Yuan J, Jiang S, Liu Y, Wang L, Liu M, et al. Asymmetrical changes of gene
702 expression, small RNAs and chromatin in two resynthesized wheat allotetraploids. Plant J.
703 2018;93(5):828–42. doi:10.1111/tpj.13805

704 Kashkush K, Feldman M, Levy AA. Gene loss, silencing and activation in a newly
705 synthesized wheat allotetraploid. Genetics. 2002;160(4):1651–9.
706 doi:10.1093/genetics/160.4.1651

707 Kaur R, Choudhury A, Chauhan S, Ghosh A, Tiwari R, Rajam MV. RNA interference and
708 crop protection against biotic stresses. Physiol Mol Biol Plants. 2021;27(10):2357–73.
709 doi:10.1007/s12298-021-01058-2

710 Kenan-Eichler M, Leshkowitz D, Tal L, Noor E, Melamed-Bessudo C, Feldman M, et al.
711 Wheat hybridization and polyploidization results in deregulation of small RNAs. Genetics.
712 2011;188(2):263–72. doi:10.1534/genetics.111.128348

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

716 Lee HS, Chen ZJ. Protein-coding genes are epigenetically regulated in Arabidopsis
717 polyploids. Proc Natl Acad Sci U S A. 2001;98(12):6753–8. doi:10.1073/pnas.121064898

718 Lee SC, Ernst E, Berube B, Borges F, Parent J-S, Ledon P, et al. Arabidopsis
719 retrotransposon virus-like particles and their regulation by epigenetically activated small
720 RNA. Genome Res. 2020;30:576–88. doi:10.1101/gr.259044.119

721 Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning
722 sequence reads to genomic features. Bioinformatics. 2014;30(7):923–30.
723 doi:10.1093/bioinformatics/btt656

724 Liu C, Wang YG. Does one subgenome become dominant in the formation and evolution of
725 a polyploid? Ann Bot. 2023;131(1):11–6. doi:10.1093/aob/mcac024

726 Mao Y, Xu J, Wang Q, Li G, Tang X, Liu T, et al. A natural antisense transcript acts as a
727 negative regulator for the maize drought stress response gene ZmNAC48. J Exp Bot.
728 2021;72(7):2790–806. doi:10.1093/jxb/erab023

729 Marí-Ordóñez A, Marchais A, Etcheverry M, Martin A, Colot V, Voinnet O. Reconstructing de
730 novo silencing of an active plant retrotransposon. Nat Genet. 2013;45(9):1029–33.
731 doi:10.1038/ng.2703

732 Martinez Palacios P, Jacquemot M-P, Tapie M, Rousselet A, Diop M, Remoué C, et al.
733 Assessing the response of small RNA populations to allopolyploidy using resynthesized

734 Brassica napus allotetraploids. Mol Biol Evol. 2019;36(4):709–26.
735 doi:10.1093/molbev/msz007

736 McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq
737 experiments with respect to biological variation. Nucleic Acids Res. 2012;40(10):4288–97.
738 doi:10.1093/nar/gks042

739 McClintock B. The significance of responses of the genome to challenge. Science.
740 1984;226:792–801. doi:10.1126/science.15739260

741 Moss E. RNA interference: it's a small RNA world. Curr Biol. 2001;11(19):R772–5.
742 doi:10.1016/S0960-9822(01)00411-1

743 Pecinka A, Fang W, Rehmsmeier M, Levy AA, Mittelsten Scheid O. Polyploidization
744 increases meiotic recombination frequency in Arabidopsis. BMC Biol. 2011;9:24.
745 doi:10.1186/1741-7007-10-33

746 Poretti M, Praz CR, Meile L, Kälin C, Schaefer LK, Schläfli M, et al. Domestication of high-
747 copy transposons underlies the wheat small RNA response to an obligate pathogen. Mol Biol
748 Evol. 2020;37(3):839–48. doi:10.1093/molbev/msz272

749 Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features.
750 Bioinformatics. 2010;26(6):841–2. doi:10.1093/bioinformatics/btq033

751 Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a
752 next generation web server for deep-sequencing data analysis. Nucleic Acids Res.
753 2016;44:W160–5. doi:10.1093/nar/gkw257

754 Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential
755 expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–40.
756 doi:10.1093/bioinformatics/btp616

757 Shen Y, Sun S, Hua S, Shen E, Ye CY, Cai D, et al. Analysis of transcriptional and
758 epigenetic changes in hybrid vigor of allopolyploid Brassica napus uncovers key roles for
759 small RNAs. Plant J. 2017;91(5):874–93. doi:10.1111/tpj.13605

760 Tiwari R, Rajam MV. RNA- and miRNA-interference to enhance abiotic stress tolerance in
761 plants. J Plant Biochem Biotechnol. 2022;31:1–10. doi:10.1007/s13562-022-00770-9

762 Tossolini I, Mencia R, Arce AL, Manavella PA. The genome awakens: transposon-mediated
763 gene regulation. Trends Plant Sci. 2025; doi:10.1016/j.tplants.2025.02.005

764 Wang HLV, Chekanova JA. Small RNAs: essential regulators of gene expression and
765 defenses against environmental stresses in plants. Wiley Interdiscip Rev RNA. 2016;7:356–
766 81. doi:10.1002/wrna.1340

Wang M, Zhao Y, Zhang B. Efficient test and visualization of multi-set intersections. Sci Rep. 2015;5:16923. doi:10.1038/srep16923

Wolfe TM, Balao F, Trucchi E, Bachmann G, Gu W, Baar J, et al. Recurrent allopolyploidizations diversify ecophysiological traits in marsh orchids (Dactylorhiza majalis s.l.). Mol Ecol. 2023;32:4777–90. doi:10.1111/mec.17070

Xiao Y, Wang J. Understanding the regulation activities of transposons in driving the variation and evolution of polyploid plant genome. Plants. 2025;14(8):1160. doi:10.3390/plants14081160

Xie Y, Ying S, Li Z, Zhang Y, Zhu J, Zhang J, et al. Transposable element-initiated enhancer-like elements generate the subgenome-biased spike specificity of polyploid wheat. Nat Commun. 2023;14(1):7465. doi:10.1038/s41467-023-42771-9

Xiong Z, Gaeta RT, Edger PP, Cao Y, Zhao K, Zhang S, et al. Chromosome inheritance and meiotic stability in allopolyploid Brassica napus. G3 (Bethesda). 2021;11(2):jkaa011. doi:10.1093/g3journal/jkaa011

Zhan J, Meyers BC. Plant small RNAs: their biogenesis, regulatory roles, and functions. Annu Rev Plant Biol. 2024;75:21–51. doi:10.1146/annurev-arplant-070122-035226

Zhang X, Xia J, Lii YE, Barrera-Figueroa BE, Zhou X, Gao S, et al. Genome-wide analysis of plant nat-siRNAs reveals insights into their distribution, biogenesis and function. Genome Biol. 2012;13:R20. doi:10.1186/gb-2012-13-3-r20

Supplementary Tables Legends

Table S1. Sample information. Collector abbreviations are FB, Francisco Balao; MC, Mark Chase; MF, Michael Fay; MH, Mikael Hedren; and OP, Ovidiu Paun. Note: the smRNA-seq data has been collected within a common garden setup; here only the population of origin of the plants is indicated.

Table S2. Criteria for categories of genomic interactions. 4x, allotetraploid; fuc - *D. fuchsii*; inc - *D. incarnata*.

Table S3. Mapping statistics for each sample. Accession IDs ending with 'r' indicate technical replicates.

Table S4. Number of smRNA targeting peaks in differential targeting (DT) analysis. Sig, significant; TE, transposable elements.

Table S5. List of genes both differentially expressed between the allotetraploids in Wolfe et al. (2023) and differentially targeted by smRNAs (this study).

Table S6. Enrichment result for genes differentially targeted between the polyploids. N_peaks - number of peaks in GO term, N_genes - number of genes in GO term, adj_pval- adjusted p-value from topGO, Z-score - normalised value based on number of genes and the average log fold change value.

Table S7. Enrichment result for genes differentially target between the polyploids, filtered for FDR < 0.05. N_peaks− number of peaks in GO term, n_genes− number of genes in GO term, adj_pval− adjusted p−value from topGO, zscore− normalised value based on number of genes and the average log fold change value.

Table S8. Number of peaks for each defined genomic interaction (additive, transgressive or dominant towards either diploid). Regions mean: all - genome wide, genes - around annotated genes, TEs - around annotated TEs, other regions - outside annotated TEs and genes. Abbreviations in the column names are as follow: maj - number of peaks in *D. majalis*, tra - number of peaks in *D. traunsteineri*, sig. Significance level for hypergeometric test done for shared. * < 0.05, ** < 0.01, *** < 0.001.

Table S9. Unfiltered enrichment result for genes around transgressive peaks in *D. majalis*. N_peaks - number of peaks in GO term, N_genes - number of genes in GO term, adj_pval- adjusted p-value from topGO, Z-score - normalised value based on number of genes and the average log fold change value.

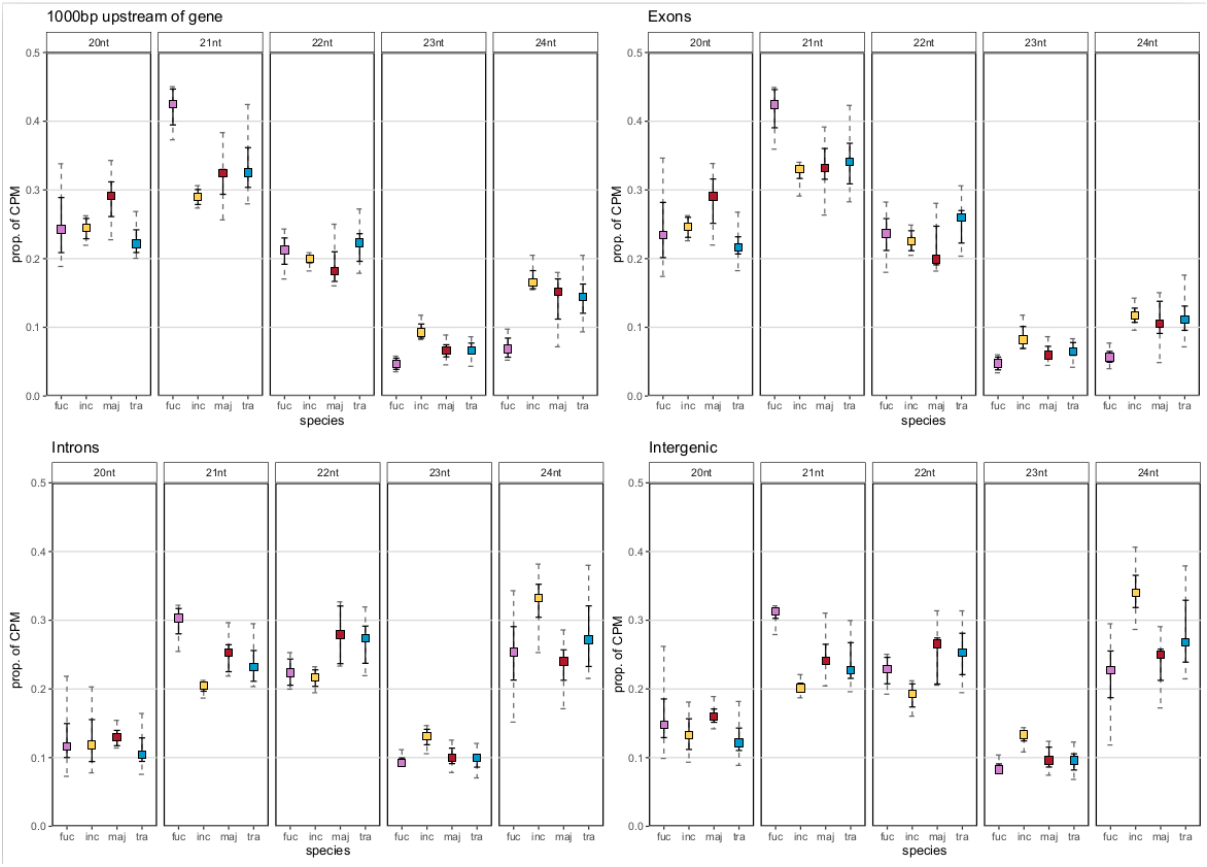
Table S10. Enrichment result for genes transgressively targeted in *D. majalis*, filtered for FDR < 0.05. N_peaks− number of peaks in GO term, n_genes− number of genes in GO term, adj_pval− adjusted p−value from topGO, zscore− normalised value based on number of genes and the average log fold change value.

Table S11. Unfiltered enrichment result for genes around transgressive peaks in *D. traunsteneri*. N_peaks - number of peaks in GO term, N_genes - number of genes in GO term, adj_pval- adjusted p-value from topGO, Z-score - normalised value based on number of genes and the average log fold change value.

Table S12. Enrichment result for genes transgressively targeted in *D. traunsteineri*, filtered for FDR < 0.05. N_peaks− number of peaks in GO term, n_genes− number of genes in GO term, adj_pval− adjusted p−value from topGO, zscore− normalised value based on number of genes and the average log fold change value.

829 **Supplementary Figures**

830



831

832 **Figure S1.** Distribution of normalised reads counts for different smRNA lengths over
833 genomic regions (promoter/1000 bp upstream, exons, introns and intergenic) in the four
834 tested species. fuc = *D. fuchsii,* inc = *D. incarnata*, maj = *D. majalis*, tra = *D. traunsteineri*.

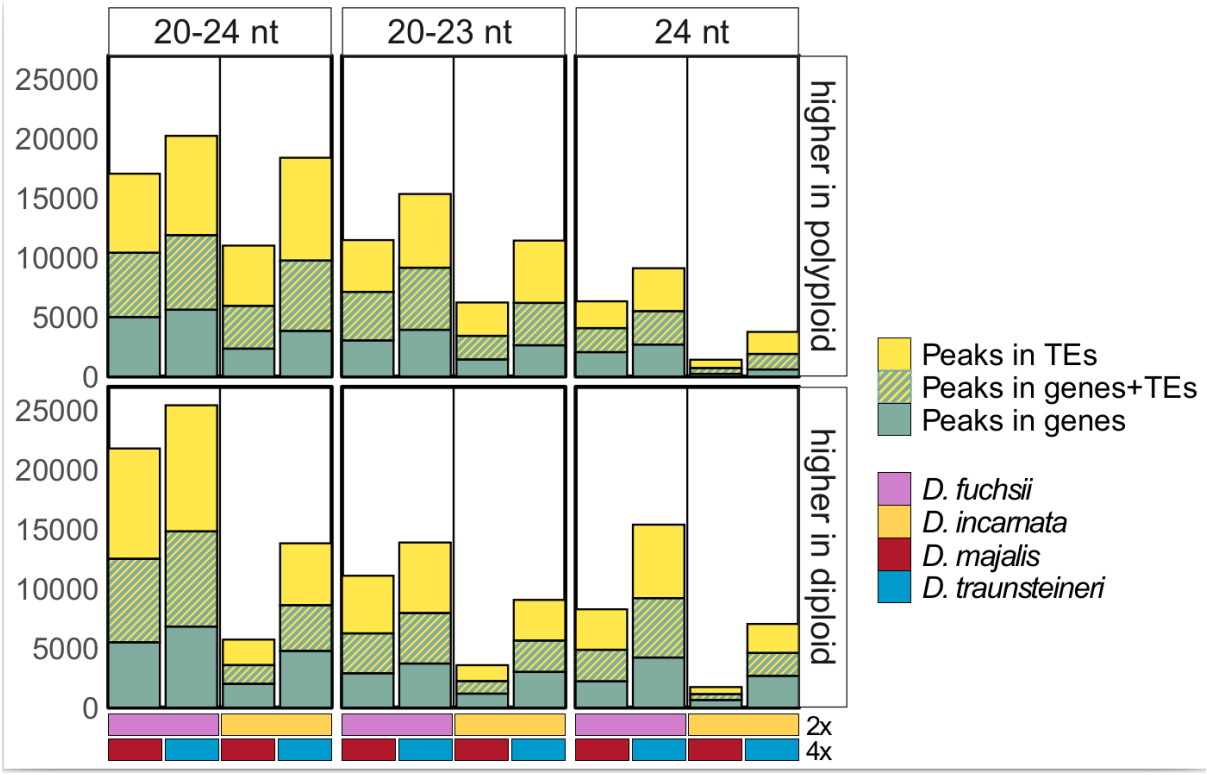


Figure S2. Number of differentially targeted peaks in genes and TEs for each comparison between a polyploid (4x) and a diploid (2x) for all three data sets, 20-24 nt 20-23 nt and 24 nt smRNAs. Top panel contains the number of peaks where the polyploid is over regulated compared to the diploid, whereas the bottom panel contains the number of peaks where the diploid is over targeted. The coloured boxed below the figure represents the comparison shown in the corresponding column.

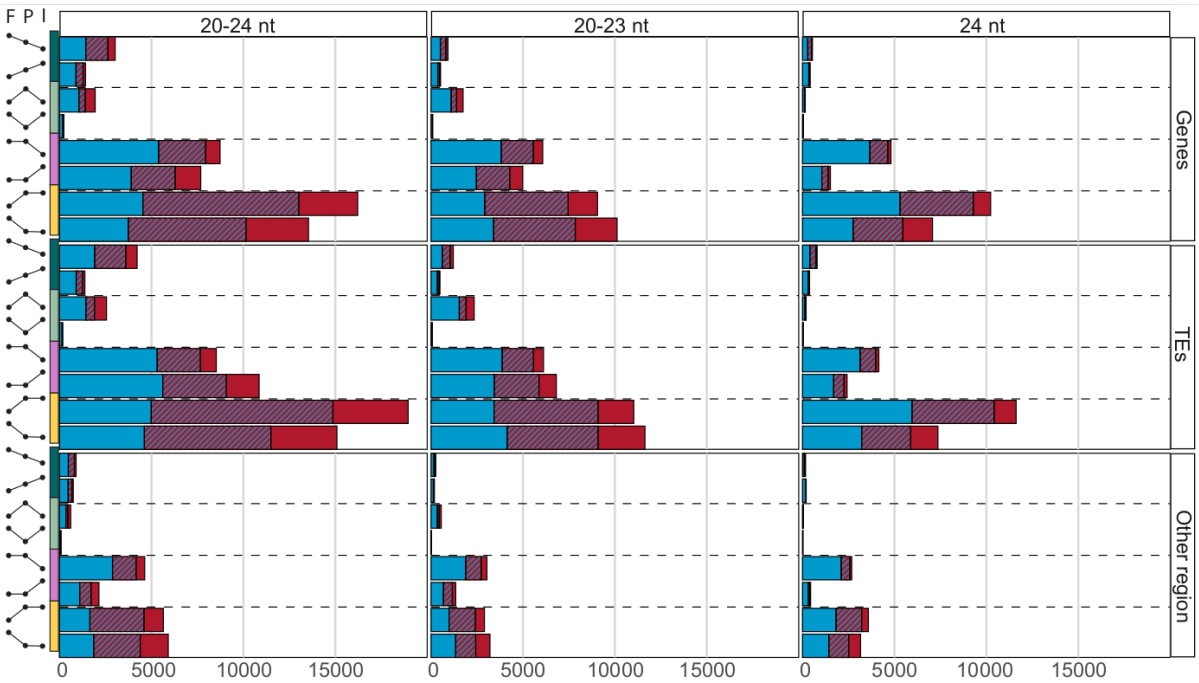


Figure S3. Number of DT peaks in all datasets 20−24 nt, 20−23 nt and 24 nt smRNAs for each genomic interaction (additive, transgressing or dominant to either diploid). Top row contains DT peaks in genes, the middle row shows DT peaks in TEs and the bottom row shows peaks found outside annotated genic and TE regions, e.g. intergentic. Colours represent, red - *D. majalis*, blue - *D. traunsteineri* and striped pattern - number of DT peaks showing the same genomic interaction in both *D. majalis* and *D. traunsteineri*. The two annotation columns along the Y−axis represent the type (coloured bars according to legend) of genomic interaction Fig. 4, while dots connected with lines show the direction. Letters above each column of dots represent **F** - *D. fuchsii*, **P** - polyploid and **I** - *D. incarnata*.

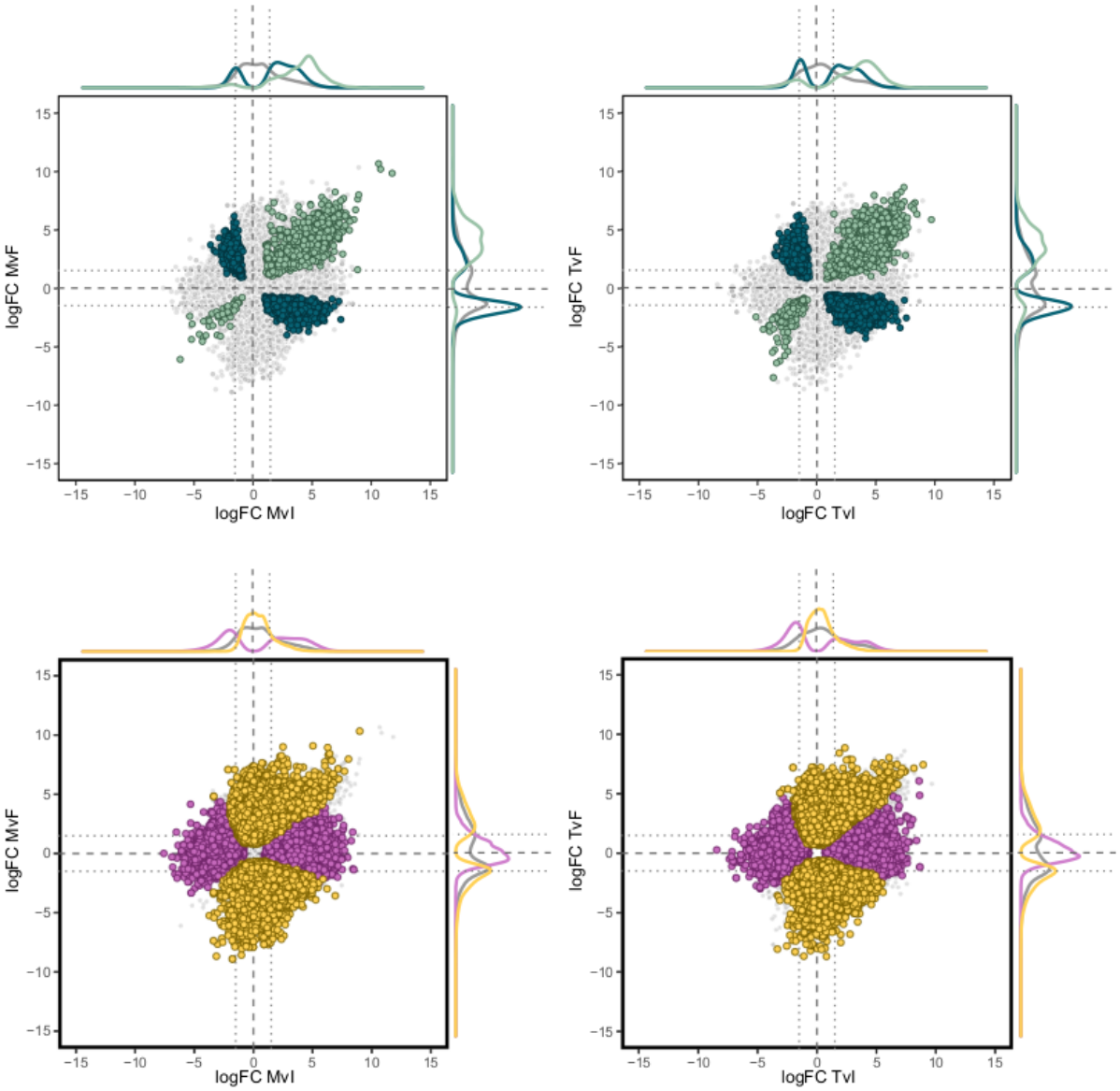


Figure S4. Patterns of genomic interactions comparing logFC values for *D. majalis* (left) and *D. traunsteineri* (right) towards either diploid, *D. fuchsii* on the Y-axis and *D. incarnata* on the X-axis for all peaks (i.e., 20−24 nt smRNAs) found in annotated genes. (a) shows with coloured symbols transgressive and additive patterns; whereas (b) is dominant to either diploid, according to the legend. Colours are according to legend of genomic interaction Fig. 4.

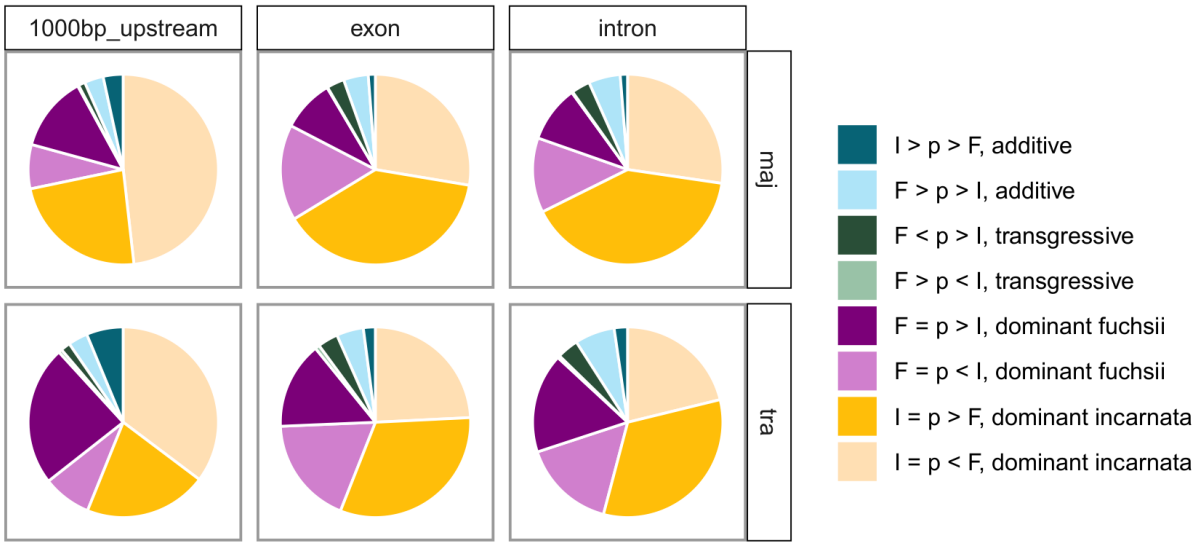


Figure S5 Proportions of peaks within each genomic interaction (additive, transgressing or dominant to either diploid) for each genomic region (1000 bp upstream/promoter, exon, intron). Colours are according to legend. Letters in legend represent **F** - *D. fuchsii*, **p** - polyploid and **I** - *D. incarnata* while symbols represent level of smRNA between the diploids and polyploid.

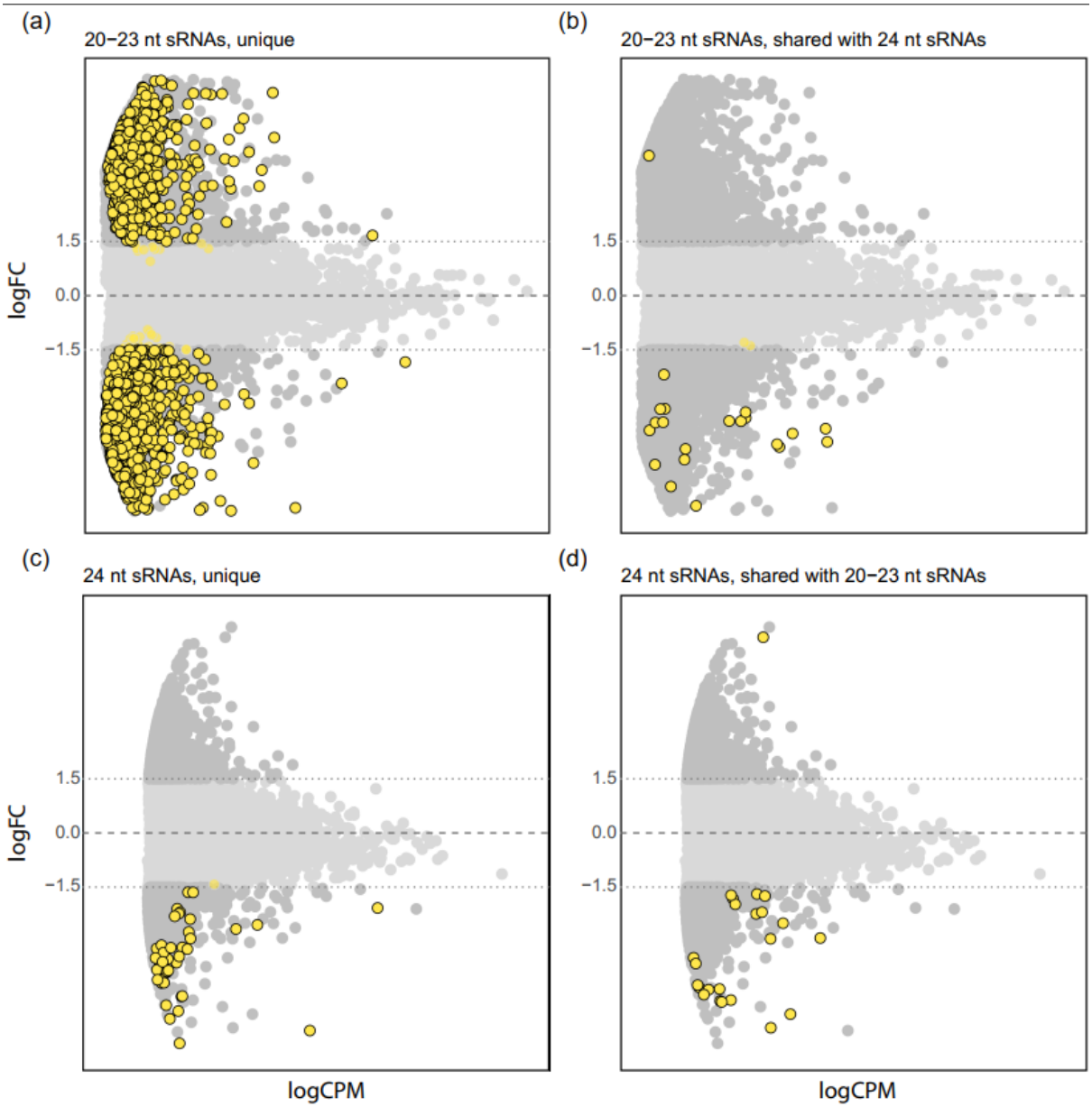


Figure S6 MA plots showing differences between *D. majalis* and *D. traunsteineri* for TEs types grouped together, the central dashed line marks 0 while the dotted lines mark logFC -1.5 and 1.5. Negative logFC values represent over-targeting in *D. majalis* as compared to *D. traunsteineri*, whereas positive logFC values indicate over-targeting in *D. traunsteineri*. (a) and (b) show 20−23 nt smRNAs while (c) and (d) show 24 nt smRNAs. (a) and (c) show DT peaks specific for each smRNA group, i.e. 20−23 nt or 24 nt, while (b) and (d) show DT peaks for both 20−23 nt and 24 nt smRNAs.

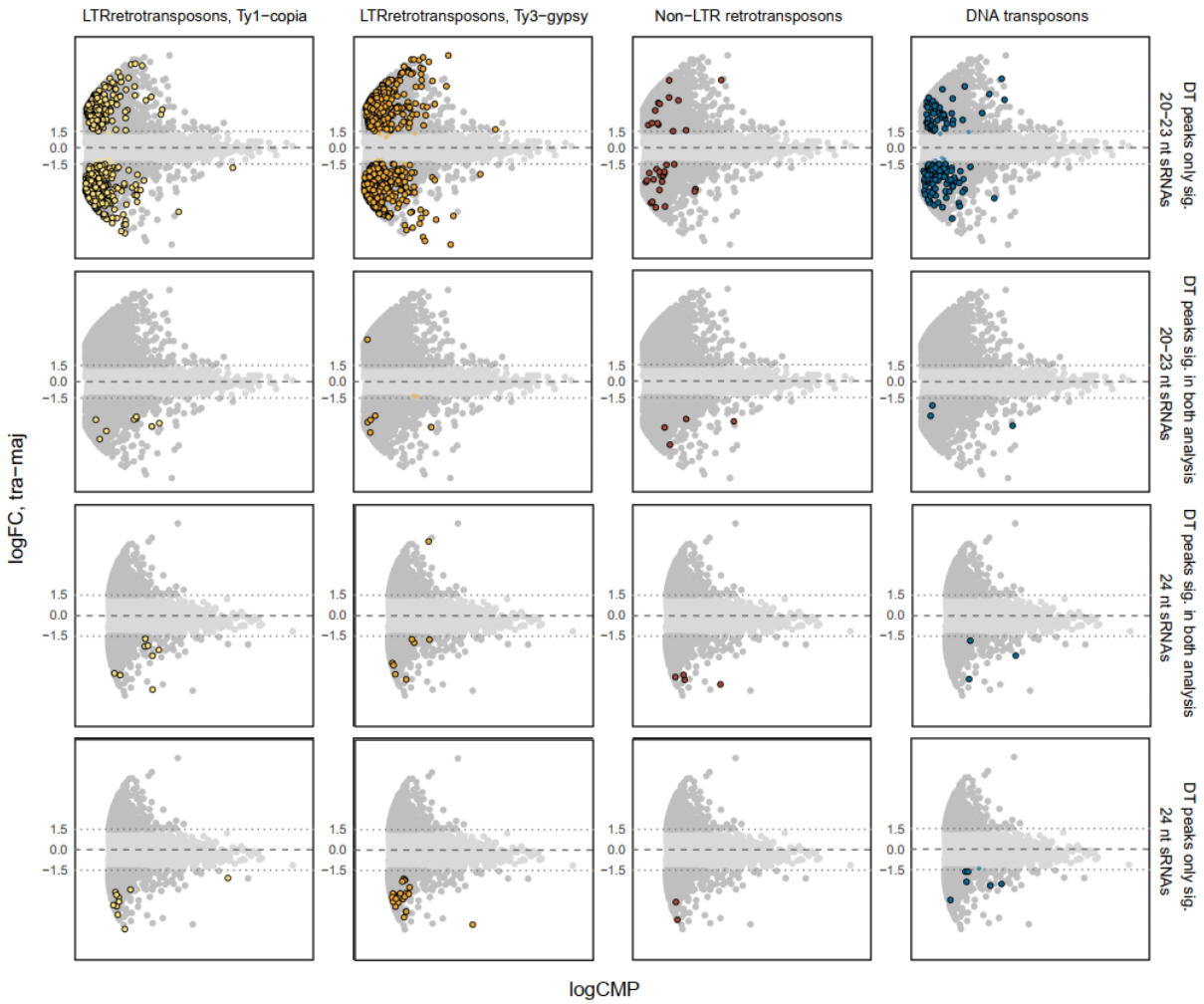


Figure S7 MA plots showing differences between *D. majalis* and *D. traunsteineri* for TEs, the central dashed line marks 0 while the dotted lines mark logFC -1.5 and 1.5. Negative logFC values represent over-targeting in *D. majalis* as compared to *D. traunsteineri*, whereas positive logFC values indicate over-targeting in *D. traunsteineri*. Colours correspond to different types of TEs, indicated by column headers. Rows correspond to smRNAs, where the 1st and 2nd row show 20−23 nt smRNAs and 3rd and 4th show 24 nt smRNAs. The 2nd and 3rd row show peaks that are DT for both 20−23 nt and 24 nt smRNAs.

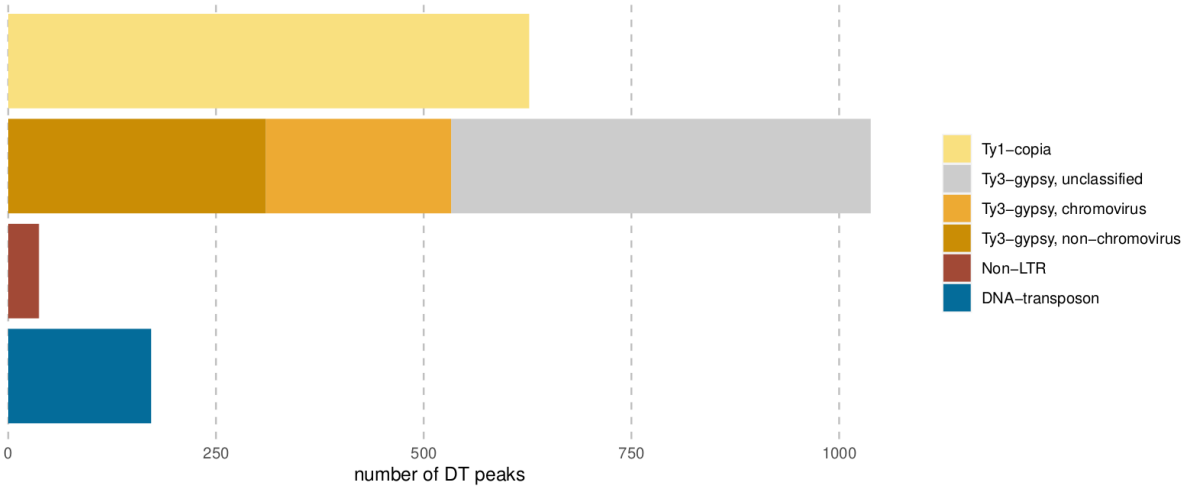


Figure S8 Count of DT peaks between *D. majalis* and *D. traunsteineri*, the coloured points of the plots in the 1st row (i.e. 20−23 nt smRNAs) in Fig. S7. Here Ty3-gypsy has been divided into the next annotation level.

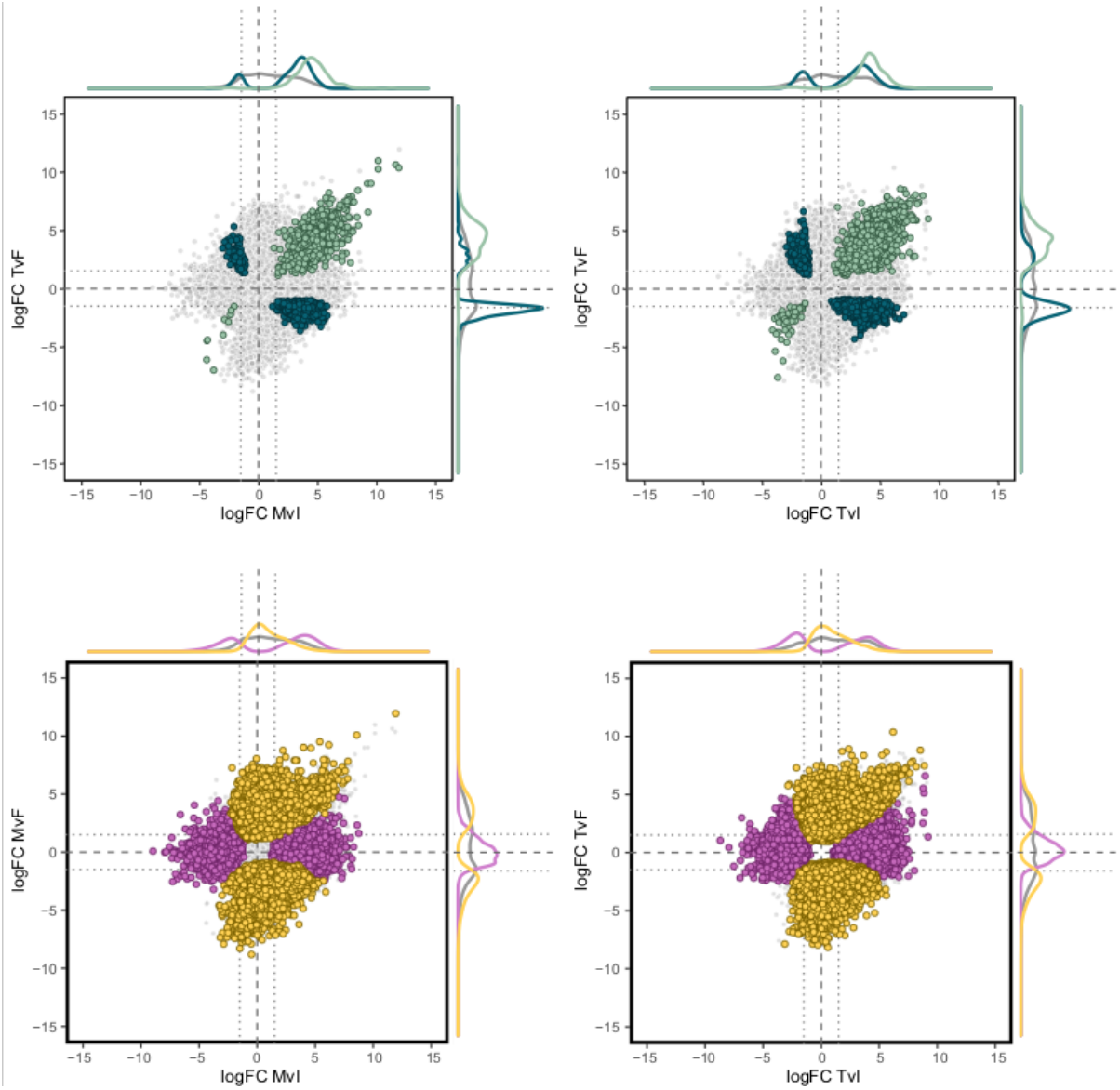


Figure S9 Patterns of genomic interactions comparing logFC values for *D. majalis* (left) and *D. traunsteineri* (right) towards either diploid, *D. fuchsii* on the Y-axis and *D. incarnata* on the X-axis for 20-23 nt smRNAs peaks found in annotated TEs. Top row shows with coloured symbols transgressive and additive patterns; whereas bottom row shows dominant to either diploid. Colours are according to legend of genomic interaction Fig. 4.

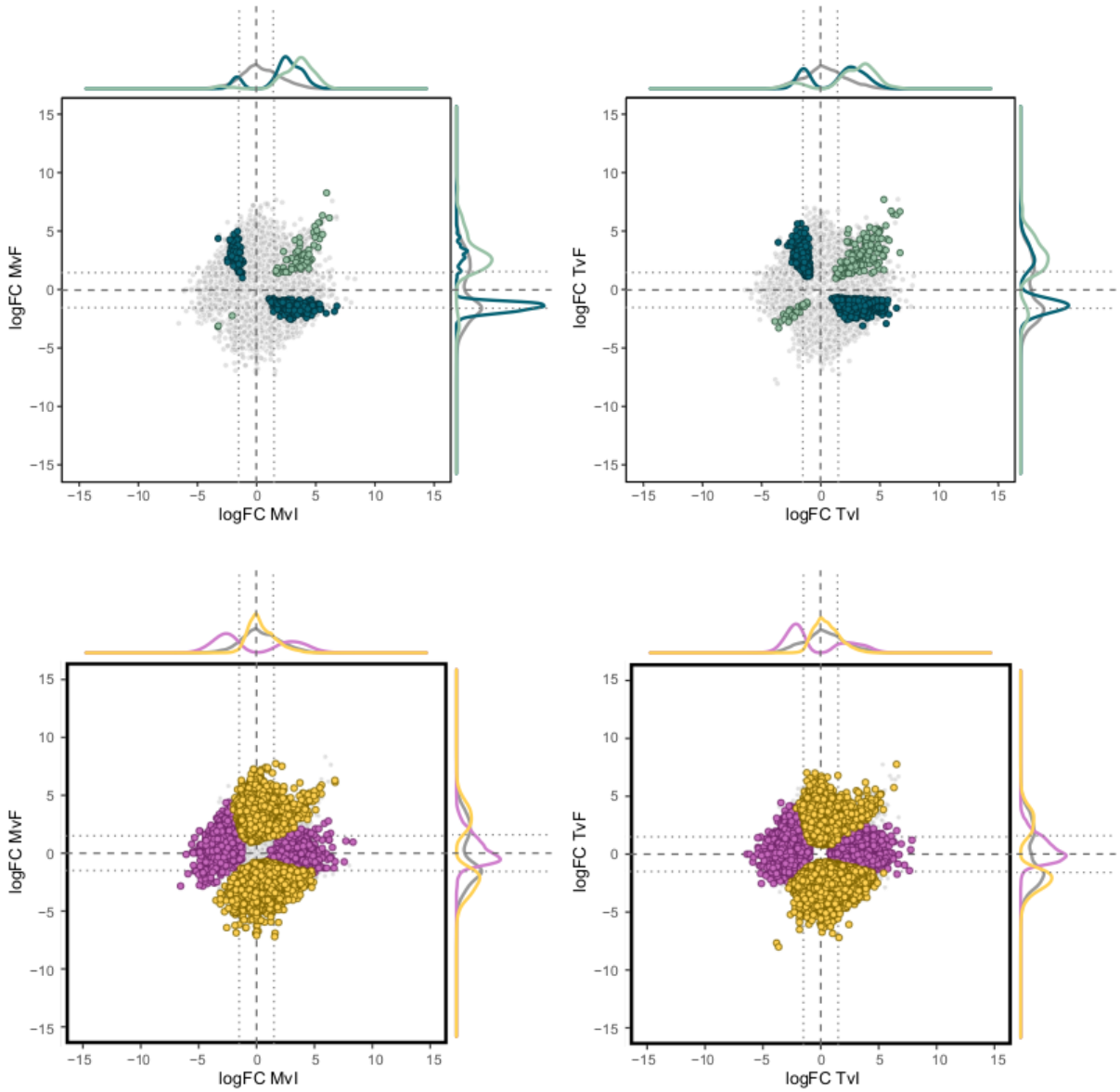


Figure S10 Patterns of genomic interactions comparing logFC values for *D. majalis* (left) and *D. traunsteineri* (right) towards either diploid, *D. fuchsii* on the Y-axis and *D. incarnata* on the X-axis for 24 nt smRNAs peaks found in annotated TEs. Top row shows with coloured symbols transgressive and additive patterns; whereas bottom row shows dominant to either diploid. Colours are according to the legend of genomic interaction Fig. 4.