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 prevalent drivers of speciation, with novel ecological properties potentially arising, among other mechanisms, through changes in gene regulation by small RNAs (smRNAs), linked to transposable element (TE) dynamics. With a common garden set-up, we comparatively investigated smRNA abundance in two young, but widely distributed, ecologically divergent sibling allotetraploid marsh orchids (*Dactylorhiza majalis* and *D. traunsteineri*) and their diploid parents. Despite independent origins, the allopolyploids appear to share a substantial portion of smRNA targeting, with transgressive smRNA targeting consistently overexpressed in both, related to key genes regulating transcription, cell division, and biotic and abiotic stress responses. TE-targeting smRNAs also display shared patterns between the sibling allopolyploids, with 20-23 nt smRNAs following the maternal and smaller genome, whereas 24 nt smRNAs targeting typically resembling the level of the paternal and larger genome. However, differences between the allopolyploids are also evident, with the older allopolyploid *D. majalis* often showing higher regulation by smRNAs, appearing more focused on fine-tuning gene copy regulation, whereas its younger sibling *D. traunsteineri* exhibits stronger non-additive expression, more prominently reflecting an apparent ongoing resolution of post-polyploidization meiotic/mitotic challenges. These findings highlight shared and species-specific smRNA dynamics, revealing how allopolyploids balance genomic instability and adaptive regulation during their evolutionary trajectories. In this system, the younger *D. traunsteineri* seems to prioritize stabilizing its genome, while the older *D. majalis* shifted towards optimizing gene expression. Together, this study emphasizes the role of smRNAs in facilitating ecological novelty and speciation during post-polyploidization evolution, providing insights into molecular mechanisms and adaptive evolution.

Keywords
allopolyploidy, *Dactylorhiza*, marsh orchids, small RNAs, transposable elements

Introduction

Allopolyploidization, i.e., the combination of whole-genome doubling and hybridization, fosters genetic variation, gene redundancy and genetic buffering, and is therefore considered a major driver of evolutionary processes and ecological diversification. However, the initial merging of two divergent genomes within the same nucleus triggers ‘genomic shock’ (McClintock 1984) and is often associated with a severe genetic bottleneck. Small noncoding RNAs (smRNAs) have emerged as critical mediators during post-polyploidization

genome and transcriptome stabilization (e.g., Ha et al., 2009; Shen et al., 2017; Cavé-Radet et al., 2020). They potentially buffer the effects of allopolyploidization by maintaining chromatin and genome stability through inherited repeat-associated small-interfering RNAs (ra-siRNAs; Shen et al., 2017), and modulating gene activity through microRNAs (miRNAs) with phenotypic implications (Ha et.al, 2009, Ghani et al., 2013, Jiang et al., 2021). Additionally, smRNAs play essential roles in responses to abiotic and biotic stress (reviewed in (reviewed in Kaur et al., 2021 and Tiwari & Rajam, 2022), facilitating adaptation to non-parental niches. This dual role in genome stabilization and stress response highlights smRNAs as pivotal players in the evolutionary success of allopolyploids.

In general, the roles of smRNAs in the genome can be divided into three main modes of regulation: (1) post-transcriptional gene silencing (PTGS), where smRNAs target mRNAs and activate/repress translation or boost cleavage and degradation, (2) transcriptional gene silencing (TGS), where smRNA mediate epigenetic modifications of loci that are sequence homologous to themselves, and (3) alternative splicing by binding to precursor mRNA (pre-mRNA) (Borges & Martienssen, 2015; Wang & Chekanova, 2016). Plant smRNAs are short non-coding RNA molecules ranging in size between 20-24 nucleotides (nt). They can be divided into two groups based on their biogenesis pathway: i) miRNAs, usually 20-22 nt long and produced from a hairpin structure, and ii) small-interfering RNAs (siRNAs), derived from double stranded RNA (dsRNA) (Borges & Martienssen, 2015; Chen et al., 2021). siRNAs can further be divided into three different types (1) repeat-associated siRNAs (ra-siRNA), also called heterochromatic siRNA (het-siRNA), usually 24 nt long and acting at the level of TGS, (2) secondary siRNA, usually 21-22nt long and acting both through TGS and PTGS and (3) natural antisense transcript siRNA (nat-siRNA) usually 21-24 nt long, and both cis- and trans-acting through PTGS (Borges & Martienssen, 2015; Chen et al., 2021).

Stress, such as environmental or whole genome duplication (WGD), can cause disruption in the TGS machinery and activate transcription of TEs leading to an accumulation of 21-22 nt repeat-derived smRNAs and a shift to a combination of TGS and PTGS regulation of TEs (Ito et al., 2011; Barber et al., 2012; Marí-Ordóñez et al., 2013; Creasey et al., 2014; Fultz et al., 2015). Further, Jiao and colleagues (2018) found a more pronounced burst of smRNAs associated with TEs in crosses from parents with different TE load compared to crosses from parents with similar TE load. Cavé-Radet and colleges (2020) compared homoploid hybrids to established allopolyploids in *Spartina* and found a weaker response of miRNAs (21-22 nt smRNAs) following hybridization suggesting a stronger effect of WGD rather than hybridization on miRNA regulation.

The orchid genus *Dactylorhiza* includes multiple independently formed sibling allotetraploid species, each likely with multiple origins involving overlapping stocks of parental populations (Brandrud et al. 2020), of which two (*D. majalis* and *D. traunsteineri*) are studied in this paper to address the role of smRNAs following allopolyploidization. The two allotetraploids share a complex history of repeated polyploidization and/or hybridization and back-crossing, but are today reasonably well separated in morphology and habitat preferences (Fig. 1a). On average, they have been estimated to have formed roughly 75,000 generations ago (*traunsteineri*) and 100,000 generations ago (*majalis*; Hawranek, 2021). Their parental species/lineages are the two diploids *D. fuchsii* (maternal) and *D. incarnata* (paternal) (Fig. 1a), that are known to differ in genome size, with *D. fuchsii* having a smaller genome (Aagard et al., 2005; Eriksson et al. 2022). A transcriptomic study of *D. fuchsii* and *D. incarnata* found large scale transcriptomic differences between the diploids, and, importantly, uncovered differences in efficiency of smRNAs control of TEs, as well as gene expression for genes involved in pathways leading to PTGS via smRNAs (Balao et al., 2017). Further, Eriksson et al. (2022) reported significant differences in both genome size and TE dynamics between the two diploids.

Distinct ecological, spatial and physiological differences have been reported for the two sibling allopolyploids (Paun et al., 2010; Paun et al., 2011; Balao et al., 2016; Wolfe et al., 2023). *Dactylorhiza majalis* is distributed across continental Europe, ranging from the Pyrenees to Scandinavia, with a broader ecological niche and less genetic differentiation than *D. traunsteineri* which has its distribution in the Alps, Scandinavia and Britain. Wolfe and colleagues (2023) reported physiological differences regarding aspects of photosynthesis and nutrient transport consistent with the divergent ecological preference of the two allotetraploids. These appear to have helped the two allotetraploids adapt to different soil chemistry conditions, with *D. traunsteineri* preferring soils depauperate in macro- and micronutrients. In stark contrast, recent analysis of the genomes of several *Dactylorhiza* allopolyploids found little difference between *D. majalis* and *D. traunsteineri* both in genome size and TE activity suggesting a strong control of TEs following recurrent allopolyploidizations (Eriksson et al. 2022).

Previous studies on the effect of smRNAs following allopolyploidization have mostly been focusing on very early stages and in resynthesized polyploids. To our knowledge only Cavé-Radet and colleagues (2019) has investigated smRNAs in allopolyploids from natural populations, although still in early stages of allopolyploid formation (*Spartina x townsendii* F1 hybrids and *S. anglica* formed ca 170 years ago). In this paper we aim to comparatively investigate smRNAs targeting in the two relatively young, but established, sibling allopolyploids *D. traunsteineri* and *D. majalis*, and its role in transcriptional and post-

transcriptional regulation of genes and TEs in this lineages. We mainly ask: (1) How does smRNA abundance in the allopolyploids relate to the parental smRNA expression levels (i.e., translated into dominance, transgressive, additive patterns), and how reproducible is their activity when comparing the independently-originated allopolyploids? (2) How actively are TEs targeted by smRNAs and are there differences between the sibling allotetraploids? (3) Are there potential links between smRNA targeting of gene expression and the divergent eco-physiological features of the two allopolyploids?

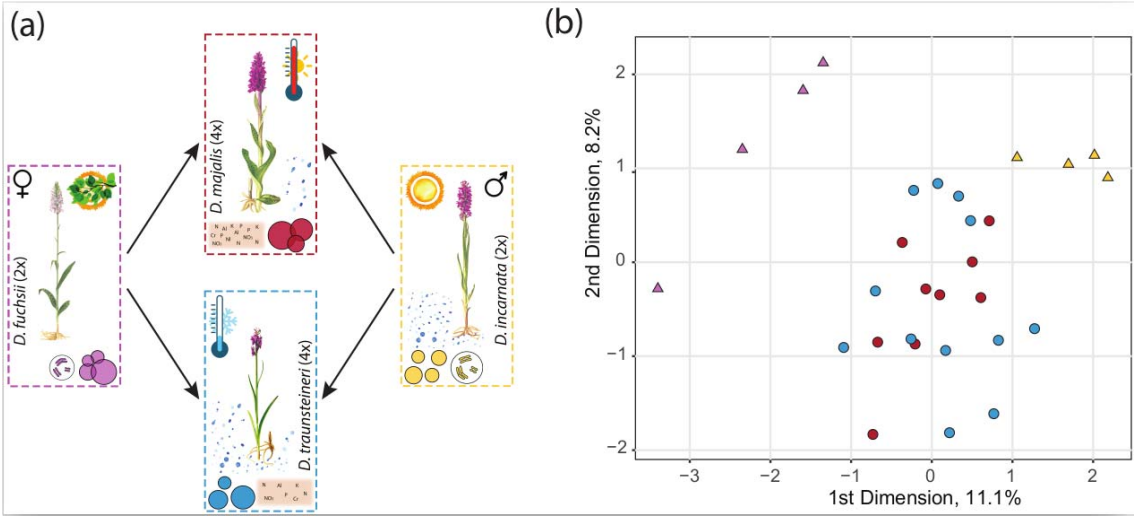

Figure 1. Two diploids (*D. fuchsii* and *D. incarnata*) and the sibling tetraploids (*D majalis* and *D. traunsteineri*) are studied in this paper. (a) Illustration depicting differences between the species. Circles and their spatial distribution towards one another reflect population densities. The cells indicate difference in genome size between the diploids and gender symbols indicate the direction of the cross. Other symbols represent broad scale ecological preferences. (b) Multidimensional scaling plot of smRNA count data with colours representing the species frames in (a) while the different shapes indicate triangles - diploid or circles - polyploid.

Materials and Methods

Sampling

Adult specimens of each species of interest were transplanted from several localities across Europe to a common garden in Vienna, Austria. *Dactylorhiza* includes perennial plants that produce a new tuber yearly, which supports growth in the following year. Two years after transplantation, in the morning of 14 May 2014, leaf tissue at a similar developmental stage for all accessions was fixed in RNAlater (Sigma) and stored at -80°C until further processing. Our final sampling included four accessions of each diploid (*D. incarnata* and *D. fuchsii*), together with nine (*D. majalis*) and twelve (*D. traunsteineri*) individuals of each allotetraploid

(Table S11). With few exceptions, each accession of each species originated from different localities across Europe. The diploid smRNA data has been previously investigated by Balao et al. (2017). Independent RNA isolations from the same tissue fixations of *D. majalis* and *D. traunsteineri* were used for RNA-seq analyses by Wolfe et al. (2021).

Library preparations and sequencing

Small RNA extraction, library preparation and sequencing for this experiment was previously described for diploid investigations (Balao et al. 2017). In short, isolation of smRNA was performed with the mirVana miRNA Isolation Kit (Life Technologies). The smRNA-enriched RNA extracts were further purified by gel size selection on 15% TBE-Urea pre-casted gels (Life Technologies). The small RNA libraries were prepared with the NEBNext Multiplex Small RNA Library Prep Set for Illumina (NEB). Illumina sequencing as 50 bp single-end reads was performed at the Vienna BioCenter Core Facilities (VBCF; https://www.viennabiocenter.org/).

Read processing and quantifying

The libraries were demultiplexed using BamIndexDecoder v.1.03 (available from http://wtsi-npg.github.io/illumina2bam/#BamIndexDecoder). The raw reads were quality filtered, trimmed from adapters used by the NEB kit and size selected for the range 20-24 nt using CLC Genomics Workbench v.8.0 (QIAGEN). The filtered reads were mapped to the available reference genome of *D. incarnata* v.1.0 (Wolfe et al., 2023) using STAR v.2.7.3a (Dobin et al., 2013), with the settings --alignIntronMax 1 and --outFilterMismatchNoverLmax 0.05. A single pass alignment approach was used, retaining further only multi-mapping reads under a threshold of 100 (--outFilterMulitmapNmax 100). A structural annotation-free approach was used for mapping, in order to also identify all regions in the genome (including non-genic) being differentially targeted (DT) by smRNA. After sorting the bam files by reference coordinates with Samtools v.1.10 (Danacek et al., 2011), read coverage was assessed for each individual along 100 bp windows of the genome using bamCoverage from deepTools2 v.3.5.1 (Ramirez et al., 2016). All windows with a minimum coverage of ten in at least one sample were retained in a bed file; neighbouring windows were merged into continuous regions using bedtools merge v.2.29.2 (Quinlan & Hall, 2010). The bed file containing retained regions across all individuals was then converted into a gff3 file with custom python script. Lastly, a table of counts was obtained running featureCounts from the Subread v.2.0.0 package (Liao et al., 2014), accounting for multiple mappers fractionally (-M --fraction).

Differential smRNA targeting analysis

Differently targeted regions were identified using R v.4.1.1 and the Bioconductor package EdgeR v.3.34.0 (Robinson et al., 2010; McCarthy et al., 2012). The count matrix was filtered to retain regions with a minimum smRNA mapping of one count per million (cpm) for a minimum of 25% of the number of samples in a group and a minimum of two samples. To account for differences in library sizes the data was normalised using the weighted 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 peaks only passing an FDR value of 0.05 were kept for further reporting.

Analyses targeting genomic regions around transposable elements were performed for reads with lengths 20-23 nt, and separately for 24 nt reads (i.e., potentially heterochromatic smRNAs, hetsiRNAs, Borges & 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. (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. (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 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 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 & 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.

Results

smRNAs and their genomic targets

Sequencing produced an average 14.0 M (±5.4 M SD) reads. After quality filtering and size selection we retained across all samples an average of 6.0 M (±3.1 M SD) reads of 20 to 24 nt (Supplementary Table S1). An average 82.9% (±4% SD) of the 20 to 24 nt reads were successfully mapped to the *D. incarnata* reference genome across all samples, with no observed mapping bias across the different species (Supplementary Table S1). The distribution of reads by length matched the expectation for smRNAs (Borges & Martienssen, 2015), showing modes at 21-22 and 24 nt across all species (Fig. 2a). In general there were higher cpm amounts of 20-22 nt smRNAs in *D. fuchsii* compared to *D. incarnata*, although only the 21 nt smRNAs were significantly different between the two diploids (Fig 2a-c). The opposite was true for the 23-24 nt smRNAs, where *D. incarnata* had higher cpm values compared to *D. fuchsii* (both the 23 nt and the 24 nt being significant) (Fig. 2a). The allotetraploids appear much more similar to each other, against the diploid comparisons (Fig. 2a). For the 21 nt smRNAs both allotetraploids have a significantly higher cpm than the paternal species, *D. incarnata*, but significantly less than the maternal *D. fuchsii* (Fig. 2a).

A large proportion of the 20-22 nt smRNAs mapped to exons, while the 24 nt mapped predominantly to TEs and intergenic regions (Fig. 2b-c, Supplementary Fig. S1). Our annotation-free approach defined 2,454,439 putative peaks of smRNA targeting across the reference genome, out of which, depending on the comparison, between 102,222 and 370,440 peaks were retained after filtering as showing higher than one cpm in at least two samples and 25% of the number of samples in a group. A multidimensional scaling (MDS) plot based on all 20-24 nt smRNAs reads distributed across 212,434 peaks clear separates the two diploids (*D. fuchsii* and *D. incarnata*), and the diploids from the allotetraploids, while the allotetraploids overlapped in the middle (Fig. 1b). However, within *D. traunsteineri* a geographical clustering was apparent.

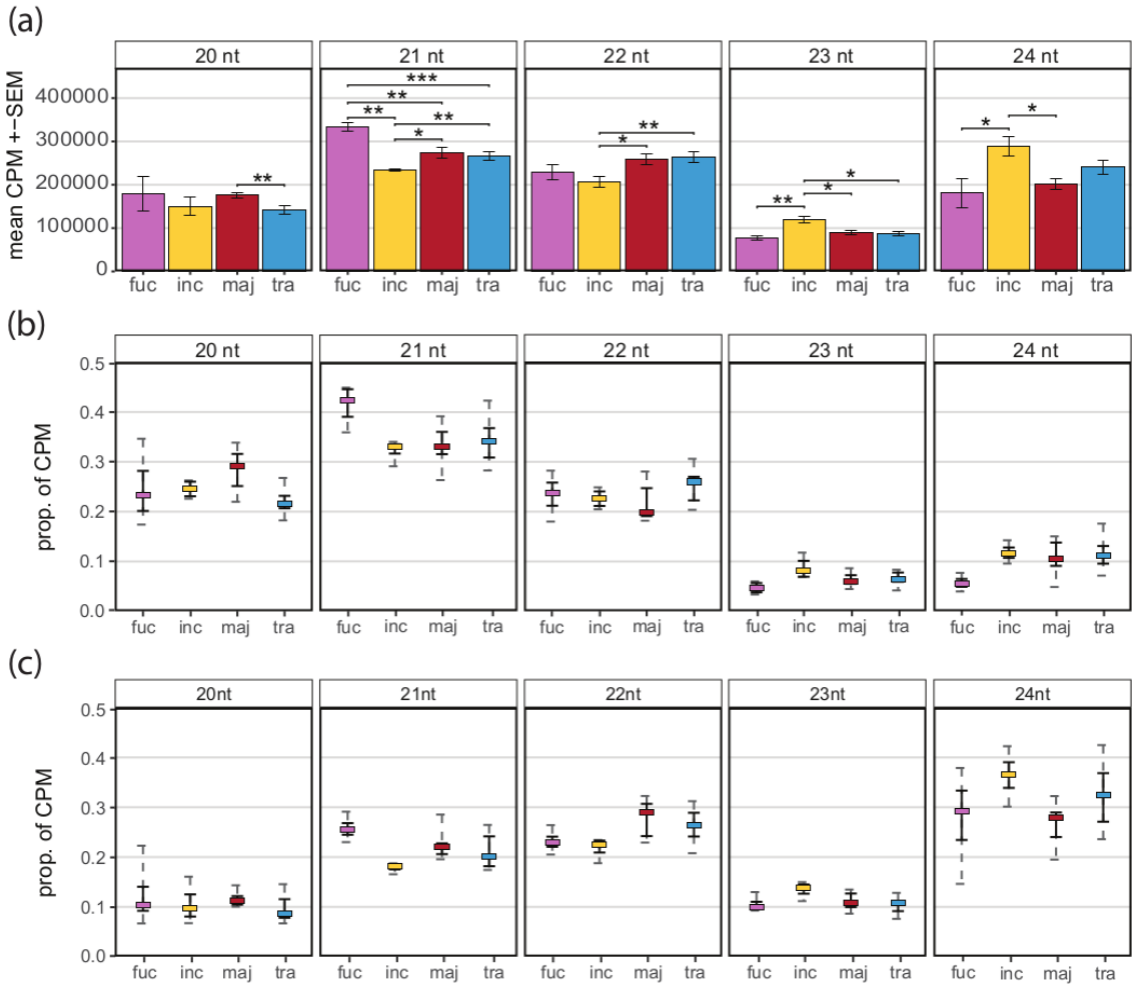

Figure 2. smRNA reads for each species divided into length groups. **(a)** shows the mean count per million (CPM) read count with error bars representing standard error of the mean. Horizontal lines indicate the significant differences between species with stars representing thresholds of * - 0.05, **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

smRNA activity in ecologically-divergent allopolyploids

We first compared the two ecologically-divergent, sibling allotetraploids across the 20 to 24 nt reads together, retaining 152,503 peaks passing the cmp filtering (Supplementary Table S2). Of these, we found 2,106 DT peaks between *D. traunsteineri* and *D. majalis*, of which about half (i.e., 1,016) was found associated with genes, defined as 1,000 bp upstream of gene models (hereafter called 'promoters'), or within a gene (including exons and introns), (Supplementary Table S2). Most DT smRNA peaks (i.e., 762) associated with genes were located within introns. Few peaks (i.e., 82) were found overlapping features (intron, exon, upstream) and therefore contained more than one feature (Supplementary Table S2). In all

genic registers there were at least three times more smRNA peaks that were over-targeted in the older allotetraploid *D. majalis* compared to *D. traunsteineri* (Supplementary Table S2), Fig. 3a-c). By comparing DT peaks with differentially expressed (DE) genes (Wolfe et al., 2023) we found that most genes DT by smRNAs do not show different expression between the allopolyploids, rather suggesting a putative alternative splicing or post-transcritional smRNA involvement (Fig. 3f, Supplementary Table S10).

Enrichment analyses for the genes DT by small RNAs between ecologically-divergent sibling allotetraploid marsh orchids have revealed an asymmetrical overrepresentation of GO terms with more functions higher targeted in *D. majalis* compared to *D. traunsteineri* (Fig. 5a , Supplementary Table S4, Supplementary Table S5). The great majority of enrichments were identified for genes targeted within gene bodies, and had functions associated with regulation of transcription and translation, transmembrane transport, response to biotic and abiotic stress, DNA repair, and regulation of meiosis and/or mitosis (Supplementary Table S5).

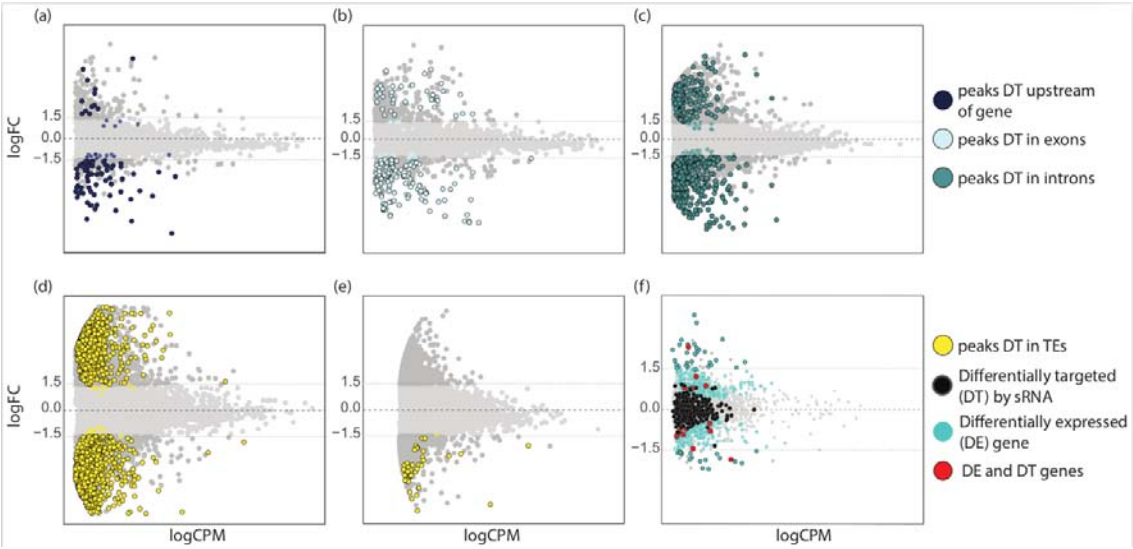

Figure 3. MA plots showing differences between *D. majalis* and *D. traunsteineri*, the central dashed line marks 0 while the dotted lines mark logFC -1.5 and 1.5. Negative logFC values represent over-targeting in *D. majalis* as compared to *D. traunsteineri*, whereas positive logFC values indicate 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 sRNAs. **(f)** An MA plot of RNA-seq data (Wolfe et al., 2021), 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 of each of the sibling allopolyploids to each of the parental diploids; after filtering, we retained 212,434 peaks for these analyses (Supplementary Table S2). Overall, both allotetraploids show similar patterns when compared to the diploids, although the younger *D. traunsteineri* shows a more deviating smRNA landscape than *D. majalis* (Supplementary Table S2, Supplementary Fig. S2). Both allotetraploids also show more differences to the maternal *D. fuchsii* than towards the paternal *D. incarnata* (Supplementary Table S2, Supplementary Fig. S2). We further characterised DT smRNAs peaks into four groups of genomic interactions: additive, transgressive, and dominant to either of the diploids. The majority of peaks remained in the class of no change observed between both diploids and the allotetraploids. Among the smRNA peaks that showed variable targeting between species, the main patterns were non-additive, with a majority towards dominance, in particular towards *D. incarnata* (Supplementary Table S3, Fig. 4a-f, Supplementary Fig. S3-S4). Although, for intergenic regions without any TEs we observed a higher portion of smRNA peaks dominant towards *D. fuchsii,* than observed for any other region (Supplementary Table S3, Fig 4d, Supplementary Fig. S3). Finally, among the transgressive patterns, in both allopolyploids most frequent were DT peaks over-targeted by smRNAs in the allopolyploid (Supplementary Table S3, Fig. 4a, Fig. 4c-f, Supplementary Fig. S3-S4).

An important general pattern over all types of genomic interaction was that a high portion of DT smRNA peaks was shared among the sibling allopolyploids (Supplementary Table S3, Fig. 4b-c, Supplementary Fig. S3-S4), despite their largely separate histories (Brandrud et al., 2020). We also investigated the distribution of each group of genomic interaction between genic regions (‘promoters’, exons, introns) and found that for DT smRNA peaks around promoters almost 75% was dominant towards *D. incarnata* in *D. majalis*, with almost 50% experiencing less smRNA targeting than *D. fuchsii* (Supplementary Fig. S5). In general, *D. traunsteineri* showed more dominance towards *D. fuchsii* across all regions (promoters, exons and introns) in comparison to the patterns observed for *D. majalis* (Supplementary Fig. S5).

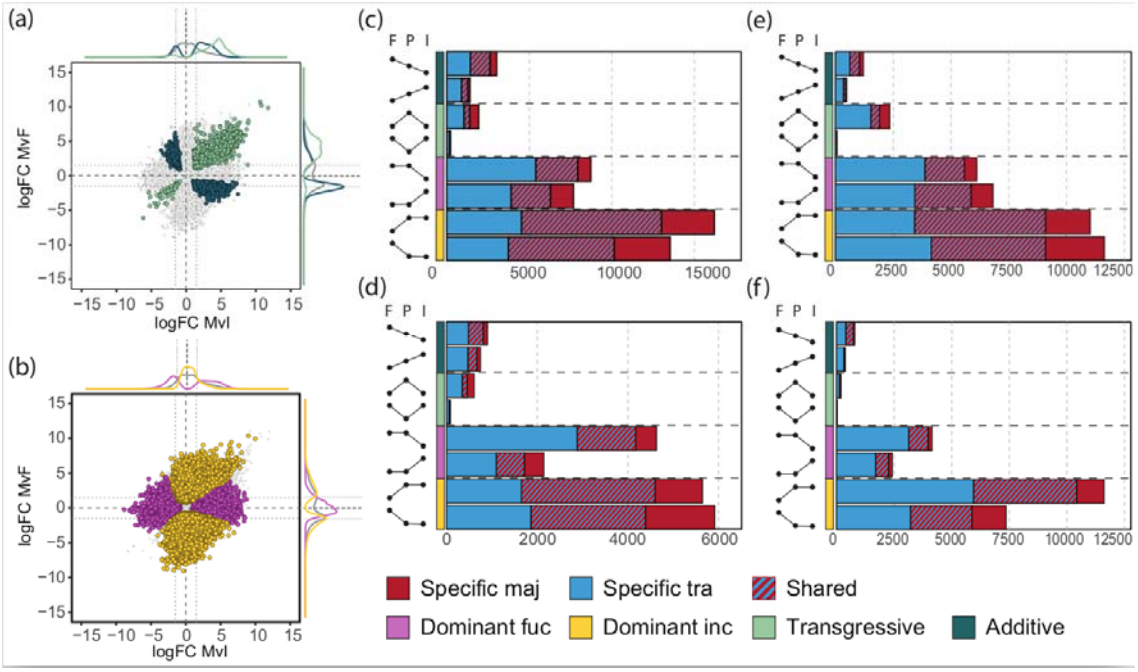

Figure 4. (a-b) Patterns of genomic interactions comparing logFC values for *D. majalis* towards either diploid, *D. fuchsii* on the Y-axis and *D. incarnata* on the X-axis for all peaks (i.e., 20-24 nt smRNAs) in annotated genes. **(a)** shows with colored symbols transgressive and additive patterns; whereas **(b)** dominant to either diploid, according to the legend. **(c-f)** Number of DT peaks in each type of genomic interaction with peaks in *D. traunsteineri* in blue and *D. majalis* in red; the striped part indicates the portion of DT peaks shared between the allotetraploids. The two annotation columns along the Y-axis represent the type (coloured bars according to legend) of genomic interaction while dots connected with lines show the direction. Letters above each column of dots represent I - *D. incarnata,* P - polyploid and F - *D. fuchsii.* **(c)** 20-24 nt smRNAs associated to genic regions, **(d)** 20-24 nt smRNA peaks outside annotated genic and TE regions, **(e)** 20-23 nt smRNAs in TEs and **(f)** 24 nt smRNAs in TEs.

For transgressively targeted genes we overall found enrichments for similar processes between the allotetraploids, such as response to stress, transport, regulation of transcription/translation and of meiosis/mitosis. Although similar processes were found, they rarely involved the same genes or the same pathways (Fig. 5b-c, Supplementary Tables S6-S9). Overall *D. traunsteineri* had an excess of enrichment for processes involved in regulating cell cycles and response to stress; these enrichments were uncovered for genes transgressively targeted by smRNAs within exons or introns (Fig. 5b, Supplementary Tables S8-S9). For *D. majalis* transgressive targeting by smRNAs affected processes involved in regulation of gene function; these enrichments were recorded for genes transgressively targeted by smRNAs only in introns (Fig. 5c, Supplementary Tables S6-S7).

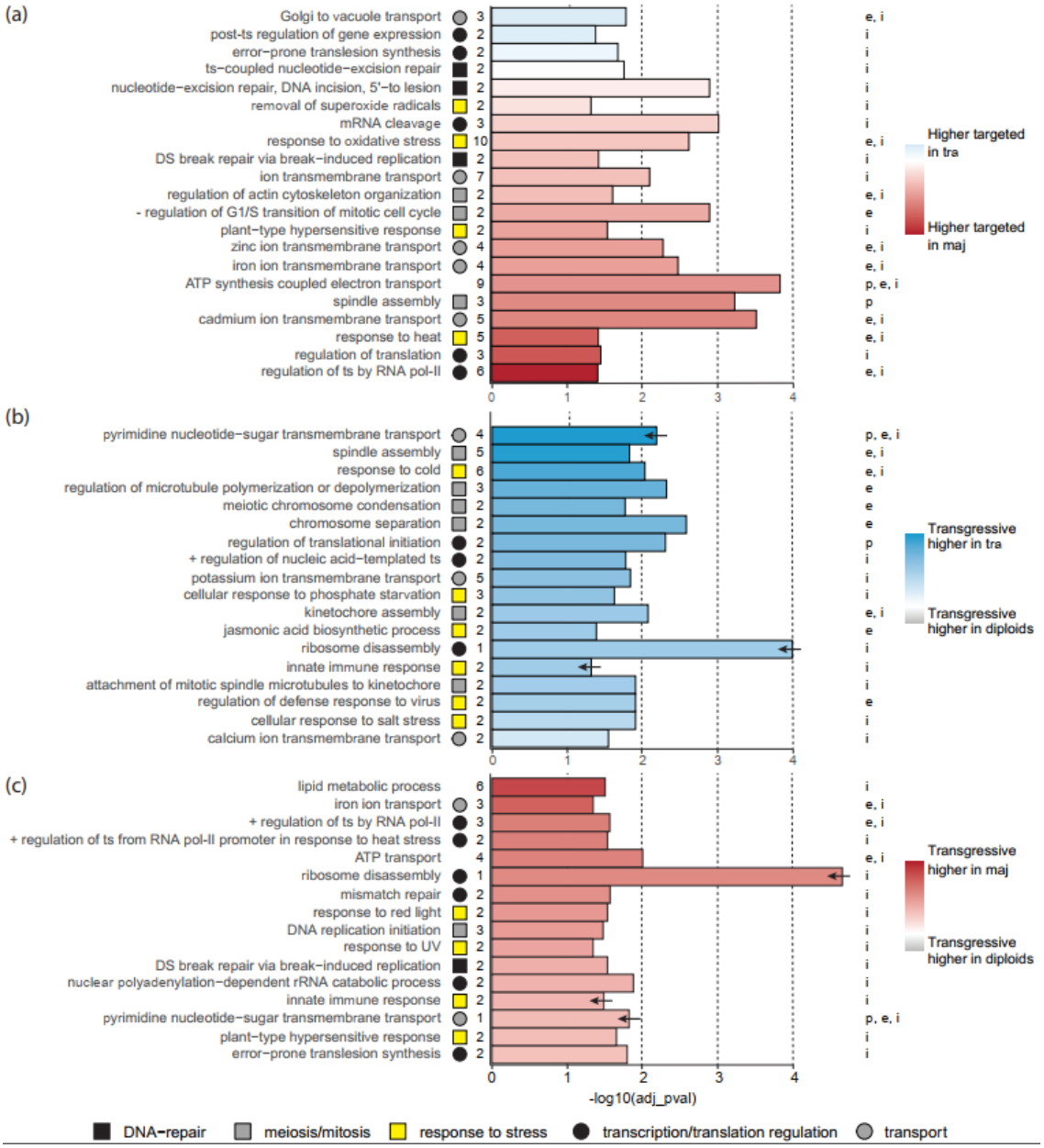

Figure 5. Summary of representative functional enrichments for **(a)** the allopolyploid comparison, **(b)** transgressive smRNA targeting of genic regions in *D. traunsteineri*, and **(c)** transgressive targeting in *D. majalis*. The size of the bars shows the adjusted significance of the enriched GO term. For the full result see Supplementary Tables S5, S7 and S9. 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.

smRNAs targeting TEs

For all analyses focusing on TEs we analysed the data in two sets (a) 24 nt smRNAs separate, since these largely represent hetsiRNAs and target TEs at a higher frequency (Fig2c, Supplementary Fig. S1) and (b) grouping 20-23 nt smRNAs together. To find DT peaks between *D. majalis* and *D. traunsteineri* we filtered the read counts for each peak by cmp and more peaks passed the filtering for the 24 nt smRNAs (315,838) compared to the 20-23 nt smRNAs (102,222) (Supplementary Table S2). In stark contrast, after the DT analysis with a generalised linear model we obtained more divergent smRNA peaks for the 20-23 nt smRNAs (3,242) than for the 24 nt smRNAs (103 peaks) (Supplementary Table S2, Fig. 3d-e). Very few TE associated peaks were found to be DT between the allotetraploids simultaneously by 20-23nt and 24nt smRNAs (Supplementary Fig. S6). Of the DT peaks in both data sets about half was associated with a TE (Supplementary Table S2). Many of the DT peaks that were associated with TEs, were also localised within introns of genes; few were also found around promoters of genes (Supplementary Table S2). In general, similarly as at the level of genic regions, among the DT peaks associated with TEs over-targeting was found more frequent in *D. majalis* compared to *D. traunsteineri*. One exception for this pattern were 20-23 nt DT peaks found within a TE located around promoters (Supplementary Table S2). The majority of TE associated DT peaks was found within LTR retrotransposons, especially Ty3-*gypsy* (Supplementary Fig. S7-S8). Out of smRNAs targeting Ty3-*gypsy* elements, slightly more were found targeting non-chromoviruses compared to chromoviruses, however, the majority was annotated as only Ty3-*gypsy* in the *D. incarnata* reference genome (Supplementary Fig. S8).

We also compared the allotetraploids towards the diploids in regard to smRNA targeting TEs, still investigating separate patterns for 20-23 nt and for 24 nt smRNAs. After cpm filtering of these datasets, we retained 139,736 and 370,440 smRNA peaks, respectively (Supplementary Table S2). Similarly to how targeting genes were analyzed, we further categorized the DT peaks into four groups of genomic interactions: additive, transgressive, and dominant to either of the diploids. In common for both datasets we recorded a significant ($p < 0.001$, hypergeometric test) amount of shared DT peaks, across categories of genomic interaction, between *D. majalis* and *D. traunsteineri* (Fig 3e-f, Supplementary Fig. S3, Fig. 4e-f, Supplementary Fig. S9-S10). Similarly as for genic regions, we found a higher frequency of dominance patterns towards either parental diploid. Regarding 20-23 nt smRNAs, slightly more smRNA peaks were found to be dominant towards the diploids less regulated by smRNAs, while the opposite was observed for 24 nt smRNAs (Fig 3e-f, Supplementary Fig. S3). Less contrast between the types of dominance (higher or lower than the other diploid) was found for 20-23 nt and a larger difference for 24 nt smRNAs (Fig

3e-f, Supplementary Fig. S3). Three patterns similar to the observed for genes were visible: Firstly, for most additive peaks *D. fuchsii* had the higher level of smRNA targeting compared to *D. incarnata* (Fig 3e-f, Supplementary Fig. S3). Secondly, for transgressive patterns the allotetraploids have in general higher levels of smRNA targeting compared to diploids (Fig 3e-f, Supplementary Fig. S3, Supplementary Fig. S9-S10). Lastly, *D. traunsteineri* showed more differences to the parental diploids, than *D. majalis* (Fig 3e-f, Supplementary Fig. S3, Supplementary Fig. S9-S10).

Discussion

In this study, we investigated the role of smRNAs to diversification following repeated allopolyploidization in the genus *Dactylorhiza* (Orchidoideae). We analysed smRNA data produced from two sibling allopolyploid and their parental diploid species grown in a common garden, in order to exclude as much as possible distinct momentary and environmental effects. Our aim was to address common and divergent patterns of (a) genomic interactions between the allopolyploids and (b) regulation of TEs and genes.

Largely consistent patterns between sibling allopolyploids and overall dominance towards paternal and larger genome

The combination of two divergent genomes in allopolyploids provides different opportunities for a neopolyploid to adjust to its duplicated genetic landscape and establish as a new species. One such possibility is through epistasis or non-additive genetic interactions, and smRNA regulation can play a key role shaping these interactions. Following allopolyploidization, non-additive effects through smRNA activity have been observed in multiple plants e.g. *Arabidopsis* (Ha et al., 2009), *Brassica napus* (Fu et al., 2016; Martinez Palacios et al., 2019) and *Spartina* (Cavé-Radet et al., 2019). Apart from the conserved smRNA, the two sibling allopolyploids in this study showed predominantly non-additive smRNA targeting, mostly dominant towards *D. incarnata*, especially regarding 24 nt TE associated smRNAs (Fig. 4, Supplementary Fig. S5). It has been illustrated that 24 nt smRNAs are mostly involved in controlling TEs, with a generally increased targeting of the subgenome with higher TE load in allopolyploids (wheat, Woodhouse et al., 2014; *Brassica rapa*, Cheng et. al., 2016). A methylation study in *Mimulus* showed that expression level dominance toward one or the other subgenome occurs immediately following allopolyploidization, and its effect increases over generations (Edger et al., 2017).

Little is known about dominant and submissive subgenomes in the two allopolyploid marsh orchids in this study. However, theory and evidence from other systems suggests that the

subgenome originating from *D. incarnata*, with its larger genome and higher TE load should become the submissive genome (Woodhouse et al., 2014, Cheng et. al., 2016, Edger et al., 2017, Gaebelein et al., 2019). Therefore, we hypothesise that the observed patterns of dominance, high degree of shared parental effects and comparatively small number of DT peaks indicate a relatively consistent regulation of TEs among the allotetraploids, possibly regulating/silencing also near-by genes in the *D. incarnata* derived subgenome. A limitation of our study however is the availability of a reference only for the paternal genome, which makes it impossible to assess homeolog specific targeting. Due to the short lengths of smRNAs and their targeting with mismatches, even when references for both genomes are available, studying subgenome effects at the level of smRNAs targeting is a difficult undertaking.

We observed more transgressive activity at peaks targeted with 20-23 nt smRNAs, compared to 24 nt smRNAs, especially when the allotetraploids are over-regulated compared to their diploid parental species (Fig. 4e-f). Genes showing transgressive smRNA targeting in the allotetraploids compared to parental diploids could be driving novel traits in allopolyploids and help them establish in a niche outside the parental range. We also observed more differentially targeted peaks around TEs between the two allopolyploids for 20-23 nt smRNAs (Fig. 3d-e). Cavé-Radet et al. (2019) reported an increase in 21 nt smRNA in F1 hybrid *Spartina x townsendii* and allopolyploid *S. anglica* compared to parental diploids. However, even if we obtained more differences in this group of smRNAs, the overall level of 21 nt smRNAs were significantly lower in the allopolyploids compared to *D. fuchsii*, while significantly higher than *D. incarnata* (Fig. 2a).

It should be noted that the observed dominance towards *D. incarnata* could be an artefact of mapping towards the *D. incarnata* reference genome. We observed an increased portion of dominance towards *D. fuchsii* in intergenic regions outside annotated TEs, although smRNAs peaks dominant towards *D. incarnata* were still the majority even in this situation. Since we obtained similar mapping success across all samples regardless of species (ca. 82.9%; Supplementary Table S1) and high levels of smRNAs mapping to genetic regions for *D. fuchsii,* a mapping bias should not play an important role for the observed pattern. However the quality of the annotation may still play a role, as to a certain degree repetitive elements of the *D. fuchsii* genome have been retained in *D. incarnata* only as degraded remnants that failed to be annotated. *Dactylorhiza incarnata* split from *D. fuchsii* roughly 5.5 Mya (Brandrud et al., 2020).

Consistent regulation of TEs by 24 nt smRNA, while more distinct regulation by 21-22 nt smRNAs among the allopolyploids

A previous study (Jiao et al., 2018) compared resynthesized wheat allopolyploids resulting from crosses between species with varying degree of divergence in the repeat content and found stronger bursts of smRNAs activity following a cross between parental species with highly differentiated TE landscapes. On the background of (*i*) the observed differences in genome sizes and TE dynamics between *D. fuchsii* and *D. incarnata* (Eriksson et al. 2022), and (*ii*) their extensive differences regarding both gene expression and smRNA targeting (Balao et al., 2017) one of our main aims was to investigate the effect of smRNAs on TEs. In general, TEs are under epigenetic control through TGS-mediated by the 23-24 nt long ra-siRNAs (Fedoroff, 2012; Woodhouse et al., 2013; Borges & Martienssen, 2015). Despite a high proportion of the 24 nt smRNAs mapping to TEs in our data we observed very few differences between the allopolyploids. In resynthesized allopolyploids it has been observed an initial increase of smRNAs in the first generation, followed by a rapid decrease over the next generations (Martinez Palacios et al., 2019). Such a quick decrease could indicate a quick response to stabilise the subgenomes following allopolyploidization by other epigenetic (e.g., DNA methylation) and genetic mechanisms. Ha et al. (2009) investigated in *Arabidopsis* the potential to buffer the 'genomic shock' often associated with allopolyploidization, and suggested that allopolyploids take advantage of smRNAs inherited from the parents to quickly achieve a high degree of genome stability. This may indicate that the allotetraploids that we study here are already too old to observe large-scale differences in TE-associated smRNAs. However, we consistently found more DT peaks that are over-targeted in the older allopolyploid *D. majalis* compared to *D. traunsteineri* (Fig. 3), suggesting that a relatively minor proportion of patterns are still being sorted by smRNA targeting after 100,000 generations following allopolyploidization.

Interestingly, we found that more smRNA differences between the allopolyploids related to TE regulation were due to the activity of 20-23 nt smRNAs. It has been found that stress can induce disruptions in the TGS regime, causing TEs to be transcribed and giving rise to 21 nt smRNAs (Ito et al., 2011; Barber et al., 2012; Marí-Ordóñez et al., 2013; Creasey et al., 2014; Lee et al., 2020). Similarly, an increase in miRNAs (typically 20-22 nt long) has been found in resynthesized allopolyploids compared to the parental species (Ha et al., 2009; Kenan-Eichler et al., 2011; Ghani et al., 2014). The same studies also show that siRNAs remain stable or even decrease in early resynthesized allopolyploids compared to parental levels. Cavé-Radet et al. (2019) extended upon this observation, reporting an increase in 21 nt ra-siRNAs in natural populations of *Spartina* allopolyploids compared to their parental species. In our data we also record a higher proportion of 21 nt smRNAs compared to the paternal species but a lower proportion towards the maternal species (Fig. 2). It is

noteworthy though, that our data regards all 21 nt smRNAs together, and not only TE associated smRNAs.

Despite Ty1-*copia* being the most abundant LTR/TE in these polyploid genomes (Eriksson et al. 2022) we observed slightly more 20-23 nt smRNAs mapping to, and possibly being derived from, Ty3-*gypsy* (Supplementary Figs. S6-S7). Most Ty3-*gypsy* related smRNAs were found in regions not annotated further than Ty3-*gypsy,* but slightly more smRNAs were found in the subgroup non-chromoviruses than chromoviruses. Interestingly, Ty3-*gypsy* elements showed different patterns between the parental diploids, with the subgroup chromoviruses being present at a higher frequency in *D. fuchsii* genome, whereas the subgroup non-chromoviruses were higher in *D. incarnata* (Eriksson et al. 2022). Of the Ty3-*gypsy* elements with further annotations, more 20-23 nt smRNAs target non-chromoviruses than chromoviruses between the allotetraploids. This observation could be an artefact due to *D. incarnata* having the higher TE load, either (a) making it more likely for TEs from this genome to be targeted by miRNAs due to their number, or (b) due to an imbalance caused by parental effects. It has been observed that following allopolyploidization in *Gossypium*, that the subgenome with higher TE load in the parental genome, the A-subgenome, lost TE copies, while the D-subgenome gained TEs copies that was originally only present on in the A-subgenome (Chen et al. 2020). These observations by Chen et al,. (2020) suggests an activation and movement from the subgenome with higher TE load to the smaller subgenome.

Implications of smRNA targeting of genic regions

Among others, polyploid formation causes (1) meiotic/mitotic difficulties, and (2) genetic redundancy. On one hand, even if allopolyploids have two full sets of divergent genomes, some chromosomes or chromosome arms may still maintain homology; hence it becomes crucial to sort homologous and homeologous to ensure correct chromosome segregation. Preferential pairing of homologs has been observed in early stages *Arabidopsis suecica* (Pecinka et al., 2011) and *Brassica napus* (Grandont et al., 2014), although high rates of homeolog related errors have also been observed in the latter system (Xiong et al., 2021). On the other hand, with multiple sets of gene copies allopolyploids must enforce genetic diploidization by suppressing redundant or unbalanced genes (Lee & Chen, 2001; Kashkush, Feldman & Levy, 2002; Koh, Soltis & Soltis, 2010). Some duplicated genes may be maintained in the genome if they serve an important function, become specialised in different tissues or developmental stage through subfunctionalization, or gain a new function through neofunctionalization (Adams et al., 2003; Blanc & Wolfe, 2004; Duarte et al., 2006; De Smet et al., 2013).

In this study we observed a high portion of, primarily, 21 nt, but also 22 nt smRNAs targeting around exonic and promoter regions. 21-22 nt smRNAs are most likely acting by PTGS since different types of smRNAS can be of this length (miRNA, nat-siRNA and secondary siRNA), although secondary siRNAs can act by TGS as well (Borges & Martienssen, 2015). Usually targeted 1,000 bp upstream of genes could be associated with promoters and cause gene silenced in cis, while smRNA targeting of exons would result in degradation or alternative splicing (Moss, 2001, Guleria et al., 2011; Borges and Martienssen, 2015). Following allopolyploidization one would expect elevated rates of gene copy regulation by smRNAs, which could explain the high degree of 21-22 nt smRNAs targeting exons and around promoter regions. However, we observe many 21 nt smRNAs targeting these regions also in the diploids, especially *D. fuchsii*. It is unclear why this is; one hypothesis could be that *D. fuchsii* experiences more stress in the common garden environment and has a smRNA response to counteract the stressing factors. Both nat-siRNAs and miRNA have been found to be involved in response to stress. Correlations between levels of nat-siRNAs and responses to various biotic and abiotic stress factors have been found for example in *Arabidopsis thaliana* and rice (Zhang et al., 2012) and specifically to drought stress in maize (Mao et al., 2021). Responses of miRNAs to different stresses (e.g. drought, salt, temperature, nutrient and radiation) in various plant species has recently been reviewed by Tiwari and Rajam (2022). Another, less likely hypothesis could be that there are gene duplication events in *D. fuchsii* and the additional gene copies have to be regulated.

In general, we observed many similar responses for transgressively smRNA target genes involved in mitosis/meiosis, regulation of gene activity, stress response and transport between the allotetraploids (Fig. 5). Even if there are many similarities in the targeted genes by smRNAs, the molecular pathways appeared different between the allotetraploids. Interestingly, many genes are over/under regulated by smRNAs in the younger *D. traunsteineri* (ca. 74k generations old, Hawranek, 2021) which are enriched for processes that are meiosis- and mitosis-related, while for the older *D. majalis* (ca. 104k generations old) we observed more enrichments for genes involved in regulation of transcription/translation. Based on our findings we hypothesise that (a) the process of sorting out the initial meiosis/mitotic difficulties may still be noticeable in *D. traunsteineri*, (b) *D. majalis* has “overcome” most of the initial meiosis/mitosis difficulties and instead deals more with gene copy number redundancy, and (c) allopolyploidzation consists of two adjustment phases towards genome stability, one fast in the first generations and one slower spanning multiple, later generations. Initial bursts of smRNAs activity followed by a quick stabilisation has been reported in many systems following polyploidization (e.g. Ha et al. 2009, Fu et al. 2016). In *Dactylorhiza* we observe (i) more differences towards the diploids for *D.*

traunsteineri, suggesting that the younger *D. traunsteineri* are still sorting out the regulatory pathways, and (ii) *D. majalis* has more upregulated smRNA peaks than *D. traunsteineri*, suggesting a more efficient regulation. Given their relatively short windows of activity, smRNAs have the potential to be an important driver of evolution, especially in an allopolyploidization context. More investigations on natural and laboratory allopolyploids are needed to further understand the role of smRNAs in post-whole genome duplication phases.

Acknowledgements

This work was financially supported by the Austrian Science Fund (FWF) through grant Y661-B16 to O.P. and through the doctoral programme (DK) grant W1225-B20 to a faculty team, including O.P. Faculty members, in particular Andreas Futschik, Robert Kofler and Christian Schlötterer, the students and the scientific advisory board of the Vienna Graduate School of Population Genetics (www.popgen-vienna.at/) are acknowledged for numerous discussions and feedback on this work. We thank Richard Bateman for help during planning and collecting, and Juliane Baar, Daniela Paun and Maria Teresa Lorenzo for their support during laboratory work. 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. Natural England and the Forestry Commission (UK), and regional county administrations in Austria, France, and Sweden are acknowledged for issuing necessary collecting permits.

Author contributions

Study conceived and designed by OP. Transplantations and sampling 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.

References

Adams, K. L., Cronn, R., Percifield, R., & Wendel, J. F. (2003). Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proceedings of the National Academy of Sciences*, *100*(8), 4649–4654. doi: 10.1073?pnas.0630618100

Alexa, A., & Rahnenfuhrer, J. (2021). topGO: Enrichment Analysis for Gene Ontology. *R Package Version v.2.44.0*.

Balao, F., Trucchi, E., Wolfe, T. M., Hao, B.-H., Lorenzo, M. T., Baar, J., … Paun, O. (2017). Adaptive sequence evolution is driven by biotic stress in a pair of orchid species (Dactylorhiza) with distinct ecological optima. *Molecular Ecology*, *26*(14), 3649–3662. doi: 10.1111/mec.14123

Barber, W. T., Zhang, W., Win, H., Varala, K. K., Dorweiler, J. E., Hudson, M. E., & Moose, S. P. (2012). Repeat associated small RNAs vary among parents and following hybridization in maize. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(26), 10444–10449. doi: 10.1073/pnas.1202073109

Blanc, G., & Wolfe, K. H. (2004). Functional divergence of duplicated genes formed by polyploidy during arabidopsis evolution. *Plant Cell*, *16*, 1679–1691. doi: 10.1105/tpc.021410

Borges, F., & Martienssen, R. A. (2015). The expanding world of small RNAs in plants. *Nature Reviews Molecular Cell Biology*, *16*(12), 727–741. doi: 10.1038/nrm4085

Brandrud, M. K., Baar, J., Lorenzo, M. T., Athanasiadis, A., Bateman, R. M., Chase, M. W., Hedrén, M., Paun, O. (2020). Phylogenomic relationships of diploids and the origins of allotetraploids in *Dactylorhiza*. *Systematic Biology*, *69*(1), 91-109. doi: 10.1093/sysbio/syz035

Carthew, R. W., & Sontheimer, E. J. (2009). Origins and Mechanisms of miRNAs and siRNAs. *Cell*, *136*(4), 642–655. doi: 10.1016/j.cell.2009.01.035

Cavé-Radet, A., Giraud, D., Lima, O., El Amrani, A., Aïnouche, M. L., & Salmon, A. (2020). Evolution of small RNA expression following hybridization and allopolyploidization: insights from *Spartina* species (Poaceae, Chloridoideae). *Plant Molecular Biology*, *102*, 55–72. doi: 10.1007/s11103-019-00931-w

Chen, X., & Rechavi, O. (2021). Plant and animal small RNA communications between cells and organisms. *Nature Reviews Molecular Cell Biology*. Nature Research. doi: 10.1038/s41580-021-00425-y

Chen, Z.J., Sreedasyam, A., Ando, A. Song, Q, De Santiago, L.M., Hulse-Kemp, A.M., ... Schmutz, J. (2020) Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement. Nature Genetics. 52, 525–533. https://doi.org/10.1038/s41588-020-0614-5

Cheng, F., Sun, C., Wu, J., Schnable, J., Woodhouse, M. R., Liang, J., … Wang, X. (2016). Epigenetic regulation of subgenome dominance following whole genome triplication in Brassica rapa. *New Phytologist*, *211*(1), 288–299. doi: 10.1111/nph.13884

Creasey, K. M., Zhai, J., Borges, F., Van Ex, F., Regulski, M., Meyers, B. C., & Martienssen, R. A. (2014). miRNAs trigger widespread epigenetically activated siRNAs from transposons in *Arabidopsis*. *Nature*, *508*, 411–415.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., Depristo, M. A., … Group, 1000 Genomes Project Analysis. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. doi: 10.1093/bioinformatics/btr330

De Smet, R., Adams, K. L., Vandepoele, K., Van Montagu, M. C. E., Maere, S., & Van de Peer, Y. (2013). Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. *Proceedings of the National Academy of Sciences*, *110*(8), 2898–2903. doi: 10.1073/pnas.1300127110

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., … Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. doi: 10.1093/bioinformatics/bts635

Duarte, J. M., Cui, L., Wall, P. K., Zhang, Q., Zhang, X., Leebens-Mack, J., … DePamphilis, C. W. (2006). Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of *Arabidopsis*. *Molecular Biology and Evolution*, *23*(2), 469–478. doi: 10.1093/molbev/msj051

Edger, P. P., Smith, R., McKain, M. R., Cooley, A. M., Vallejo-Marin, M., Yuan, Y., … Puzey, J. R. (2017). Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. *Plant Cell*, *29*(9), 2150–2167. doi: 10.1105/tpc.17.00010

Eriksson, M.C., Mandáková, T, McCann, J, Temsch, E.M., Chase, M.W., Hedrén, M, Weiss-Schneeweiss, H, & Paun, O. (2022). Repeat dynamics across timescales: a perspective from sibling allotetraploid marsh orchids (*Dactylorhiza majalis* s.l.), *Molecular Biology and Evolution*, 39(8), msac167. doi.org/10.1093/molbev/msac167

Fedoroff, N,V. (2012). Transposable elements, epigenetics, and genome evolution. *Science* 338: 758–767. doi: 10.1126/science.338.6108.758

Fu, Y., Xiao, M., Yu, H., Mason, A. S., Yin, J., Li, J., … Fu, D. (2016). Small RNA changes in synthetic *Brassica napus*. *Planta*, *244*(3), 607–622. doi: 10.1007/s00425-016-2529-z

Fultz, D., Choudury, S. G., & Slotkin, R. K. (2015). Silencing of active transposable elements in plants. *Current Opinion in Plant Biology*, *27*, 67–76. doi: 10.1016/j.pbi.2015.05.027

Gaebelein, R., Schiessl, S. V., Samans, B., Batley, J., & Mason, A. S. (2019). Inherited allelic variants and novel karyotype changes influence fertility and genome stability in *Brassica* allohexaploids. *New Phytologist*. doi: 10.1111/nph.15804

Ghani, M. A., Li, J., Rao, L., Raza, M. A., Cao, L., Yu, N., … Chen, L. (2014). The role of small RNAs in wide hybridisation and allopolyploidisation between *Brassica rapa* and *Brassica nigra*. *BMC Plant Biology*, *14*(1), 1–13. doi: 10.1186/s12870-014-0272-9

Grandont, L., Cuñado, N., Coriton, O., Huteau, V., Eber, F., Chèvre, A. M., … Jenczewski, E. (2014). Homoeologous chromosome sorting and progression of meiotic recombination in *Brassica napus*: Ploidy does matter! *Plant Cell*, *26*, 1448–1463. doi: 10.1105/tpc.114.122788

Guleria, P., Mahajan, M., Bhardwaj, J., Yadav, S. (2011). Plant small RNAs: biogenesis, mode of action and their roles in abiotic stresses. *Genomics, Proteomics & Bioinformatics*, 9 , 183–99.

Ha, M., Lu, J., Tian, L., Ramachandran, V., Kasschau, K. D., Chapman, E. J., … Chen, Z. J. (2009). Small RNAs serve as a genetic buffer against genomic shock in *Arabidopsis* interspecific hybrids and allopolyploids. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(42), 17835–17840. doi: 10.1073/pnas.0907003106

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

Ito, H., Gaubert, H., Bucher, E., Mirouze, M., Vaillant, I., & Paszkowski, J. (2011). An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature*, *472*(7341), 115–120. doi: 10.1038/nature09861

Jiang, X., Song, Q., Ye, W., & Chen, Z. J. (2021). Concerted genomic and epigenomic changes accompany stabilization of *Arabidopsis* allopolyploids. *Nature Ecology & Evolution*, *5*(10), 1382–1393. doi: 10.1038/s41559-021-01523-y

Jiao, W., Yuan, J., Jiang, S., Liu, Y., Wang, L., Liu, M., … Chen, Z. J. (2018). Asymmetrical changes of gene expression, small RNAs and chromatin in two resynthesized wheat allotetraploids. *Plant Journal*, *93*(5), 828–842. doi: 10.1111/tpj.13805

Kashkush, K., Feldman, M., & Levy, A. A. (2002). Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. *Genetics*, 160(4), 1651–1659. doi: 10.1093/genetics/160.4.1651

Kenan-Eichler, M., Leshkowitz, D., Tal, L., Noor, E., Melamed-Bessudo, C., Feldman, M., & Levy, A. A. (2011). Wheat hybridization and polyploidization results in deregulation of small RNAs. *Genetics*, *188*(2), 263–272. doi: 10.1534/genetics.111.128348

Koh, J., Soltis, P. S., & Soltis, D. E. (2010). Homeolog loss and expression changes in natural populations of the recently and repeatedly formed allotetraploid *Tragopogon mirus* (Asteraceae). *BMC Genomics*, *11*(1), 1–16. doi: 10.1186/1471-2164-11-97

Lee, H.-S., & Chen, Z. J. (2001). Protein-Coding Genes are Epigenetically Regulated in *Arabidopsis* Polyploids. *Proceedings of the National Academy of Sciences of the United States of America*, 98(12), 6753–6758.

Lee, S. C., Ernst, E., Berube, B., Borges, F., Parent, J.-S., Ledon, P., … Martienssen, R. A. (2020). *Arabidopsis* retrotransposon virus-like particles and their regulation by epigenetically activated small RNA. *Genome Research*, *30*, 576–588. doi: 10.1101/gr.259044.119

Liao, Y., Smyth, G. K., & Shi, W. (2014). Sequence analysis featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. *30*(7), 923–930. doi: 10.1093/bioinformatics/btt656

Mao, Y., Xu, J., Wang, Q., Li, G., Tang, X., Liu, T., … Lu, Y. (2021). A natural antisense transcript acts as a negative regulator for the maize drought stress response gene ZmNAC48. *Journal of Experimental Botany*, *72*(7), 2790–2806. doi: 10.1093/jxb/erab023

Marí-Ordóñez, A., Marchais, A., Etcheverry, M., Martin, A., Colot, V., & Voinnet, O. (2013). Reconstructing *de novo* silencing of an active plant retrotransposon. *Nature Genetics*, *45*(9), 1029–1039. doi: 10.1038/ng.2703

Martinez Palacios, P., Jacquemot, M.-P., Tapie, M., Rousselet, A., Diop, M., Remoué, C., … Alix, K. (2019). Assessing the Response of Small RNA Populations to Allopolyploidy Using Resynthesized *Brassica napus* Allotetraploids. *Molecular Biology and Evolution*, *36*(4), 709–726. doi: 10.1093/molbev/msz007

McCarthy, D. J., Chen, Y., & Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research*, *40*(10), 4288–4297. doi: 10.1093/nar/gks042

McClintock, B. (1984). The significance of responses of the genome to challenge. *Science*, *226*, 792–801.

Moss, E. (2001). RNA interference: it's a small RNA world. *Current Biology*, 11 (19), R772–5.

Pecinka, A., Fang, W., Rehmsmeier, M., Levy, A. A., & Mittelsten Scheid, O. (2011). Polyploidization increases meiotic recombination frequency in *Arabidopsis*. *BMC Biology*, *9*, 24. doi: 10.1186/1741-7007-10-33

Poretti, M., Praz, C. R., Meile, L., Kälin, C., Schaefer, L. K., Schläfli, M., … Bourras, S. (2020). Domestication of high-copy transposons underlays the wheat small RNA response to an obligate pathogen. *Molecular Biology and Evolution*, *37*(3), 839–848. doi: 10.1093/molbev/msz272

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. doi: 10.1093/bioinformatics/btq033

Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., … Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, *44*, W160–W165. doi: 10.1093/nar/gkw257

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. doi: 10.1093/bioinformatics/btp616

Shen, Y., Sun, S., Hua, S., Shen, E., Ye, C. Y., Cai, D., … Fan, L. (2017). Analysis of transcriptional and epigenetic changes in hybrid vigor of allopolyploid *Brassica napus* uncovers key roles for small RNAs. *Plant Journal*, *91*(5), 874–893. doi: 10.1111/tpj.13605

Tiwari, R., & Rajam, M. V. (2022). RNA- and miRNA-interference to enhance abiotic stress tolerance in plants. *Journal of Plant Biochemistry and Biotechnology*, *9*. doi: 10.1007/s13562-022-00770-9

Wang, H. L. V., & Chekanova, J. A. (2016). Small RNAs: Essential regulators of gene expression and defenses against environmental stresses in plants. *Wiley Interdisciplinary Reviews: RNA*, *7*, 356–381. doi: 10.1002/wrna.1340

Wang, M., Zhao, Y., & Zhang, B. (2015). Efficient Test and Visualization of Multi-Set Intersections. *Scientific Reports*, *5*, 16923. doi: 10.1038/srep16923

Wolfe, T. M., Balao, F., Trucchi, E., Bachmann, G., Gu, W., Baar, J., Hedrén, M., Weckwerth, W., Leitch, A. R., & Paun, O. (2023). Recurrent allopolyploidizations diversify ecophysiological traits in marsh orchids (*Dactylorhiza majalis* s.l.). *Molecular Ecology*, 32, 4777–4790. https://doi.org/10.1111/mec.17070

Xiong, Z., Gaeta, R. T., Edger, P. P., Cao, Y., Zhao, K., Zhang, S., & Chris Pires, J. (2021). Chromosome inheritance and meiotic stability in allopolyploid *Brassica napus*. *G3: Genes, Genomes, Genetics*, *11*(2). doi: 10.1093/G3JOURNAL/JKAA011

Zhang, X., Xia, J., Lii, Y. E., Barrera-Figueroa, B. E., Zhou, X., Gao, S., … Jin, H. (2012). Genome-wide analysis of plant nat-siRNAs reveals insights into their distribution, biogenesis and function. *Genome Biology*, *13*, R20. doi: 10.1186/gb-2012-13-3-r20

Supplementary Figures

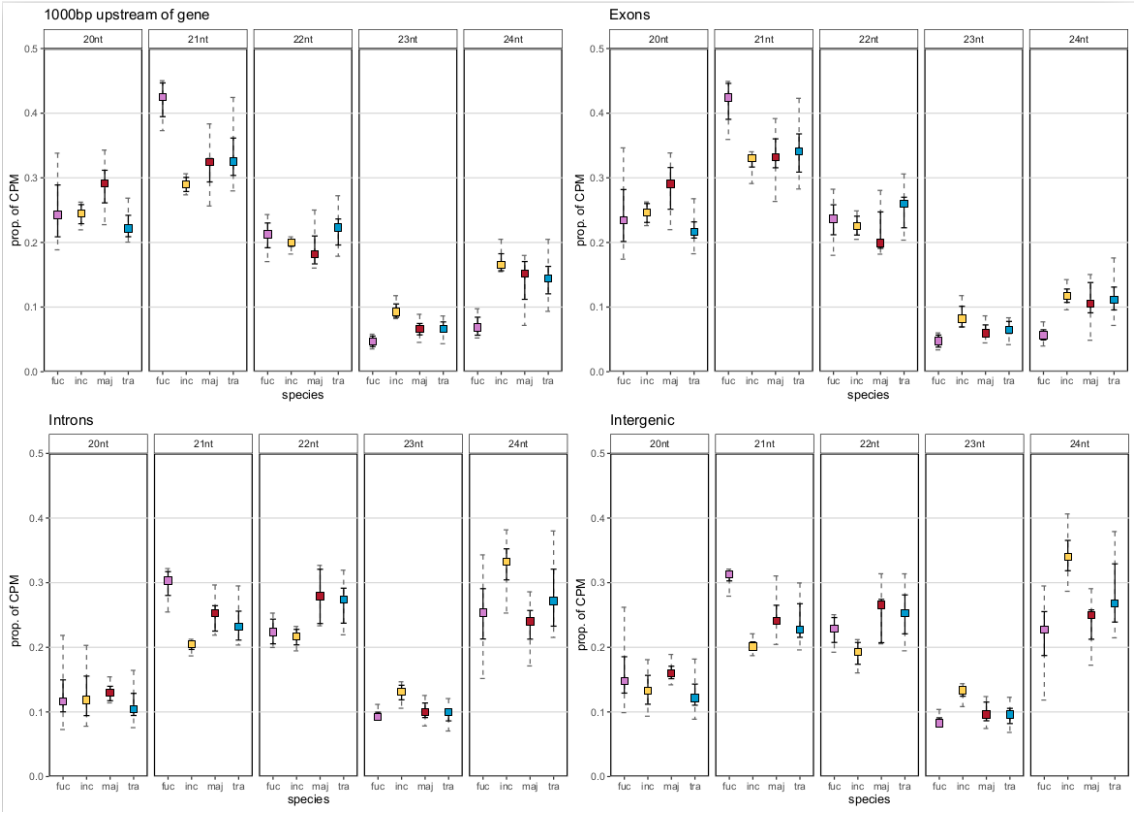

Figure S1. Distribution of normalised reads counts for different smRNA lengths over genomic regions (promoter/1000 bp upstream, exons, introns and intergenic) in the four tested species. fuc = *D. fuchsii,* inc = *D. incarnata*, maj = *D. majalis*, tra = *D. traunsteineri*.

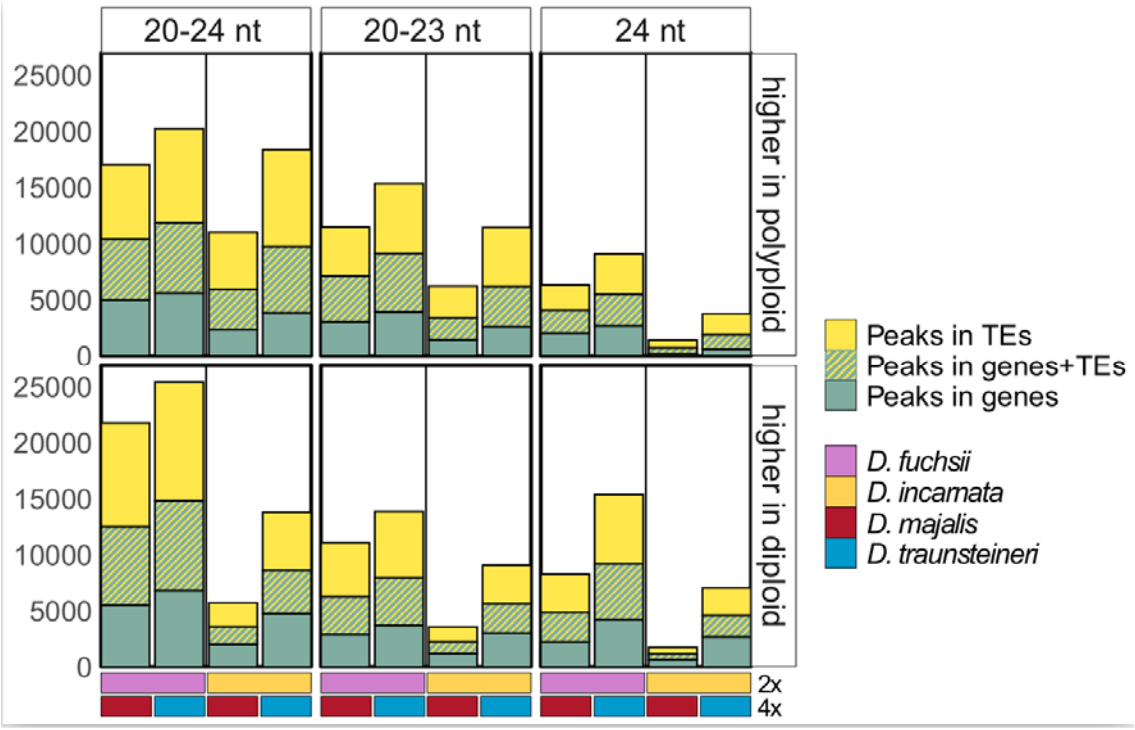

Figure S2. Number of differentially targeted peaks in genes and TEs for each comparison between a polyploid (4x) and a diploid (2x) for all three data sets, 20-24 nt 20-23 nt and 24 nt smRNAs. Top panel contains the number of peaks where the polyploid is over regulated compared to the diploid, whereas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

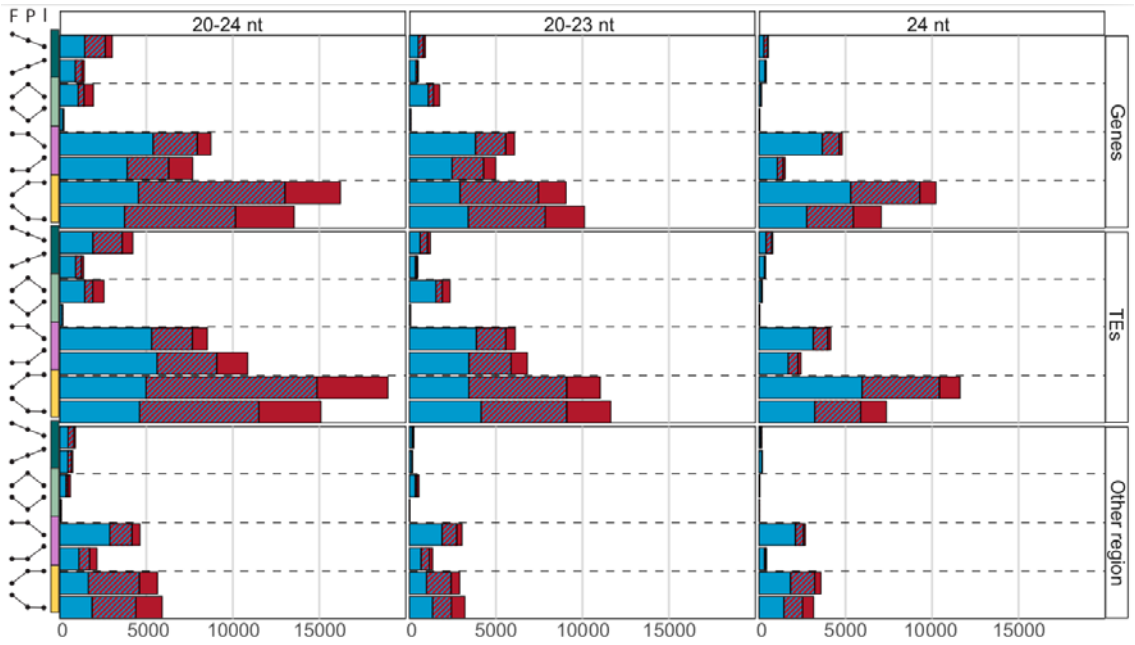

Figure S3. Number of DT peaks in all datasets 20−24 nt, 20−23 nt and 24 nt smRNAs for each genomic interaction (additive, transgressing or dominant to either diploid). Top row contains DT peaks in genes, the middle row shows DT peaks in TEs and the bottom row shows peaks found outside annotated genic 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*.

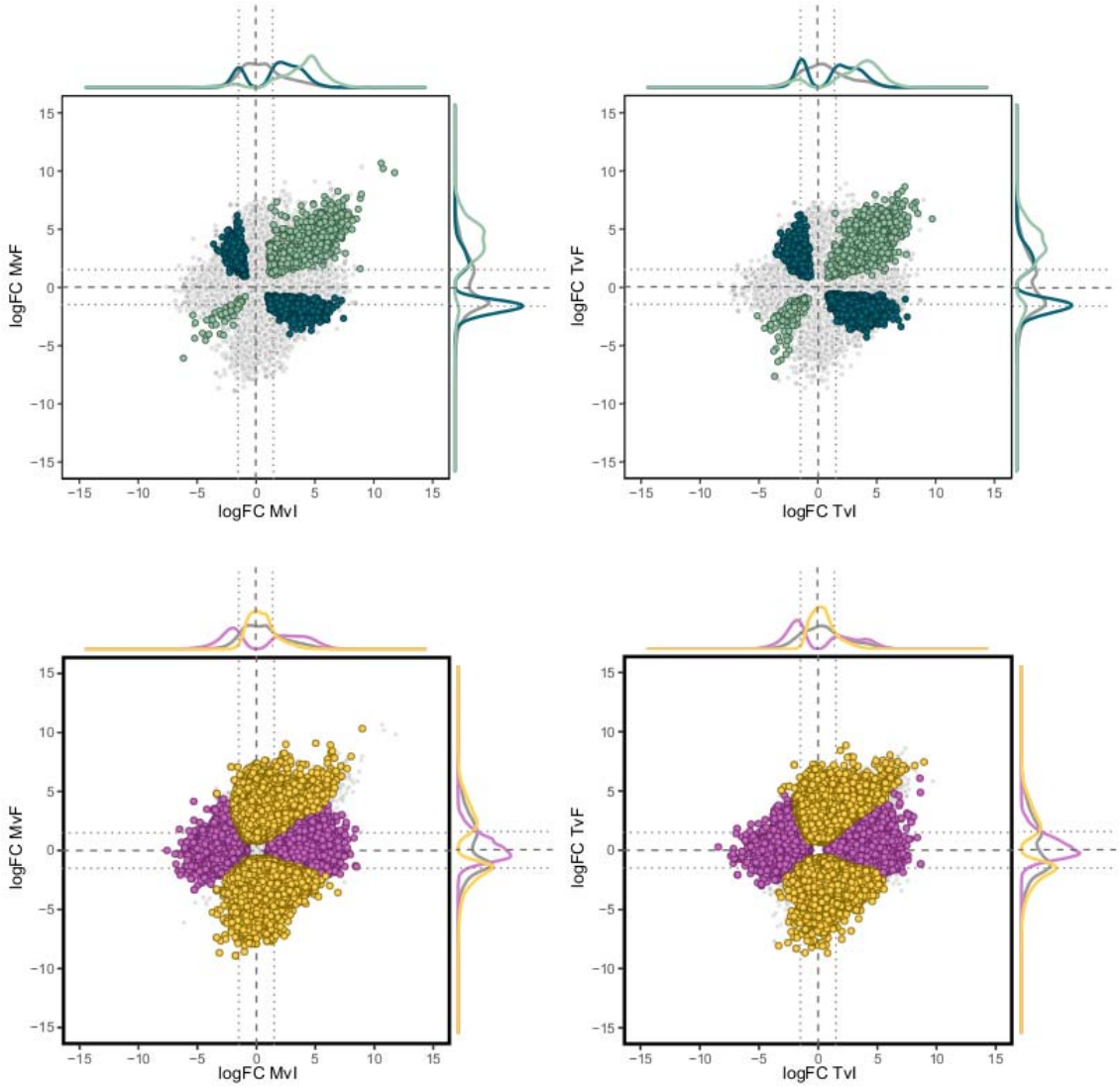

Figure S4. Patterns of genomic interactions comparing logFC values for *D. majalis* (left) and *D. traunsteineri* (right) towards either diploid, *D. fuchsii* on the Y-axis and *D. incarnata* on the X-axis for all peaks (i.e., 20−24 nt smRNAs) found in annotated genes. (a) shows with coloured symbols 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.

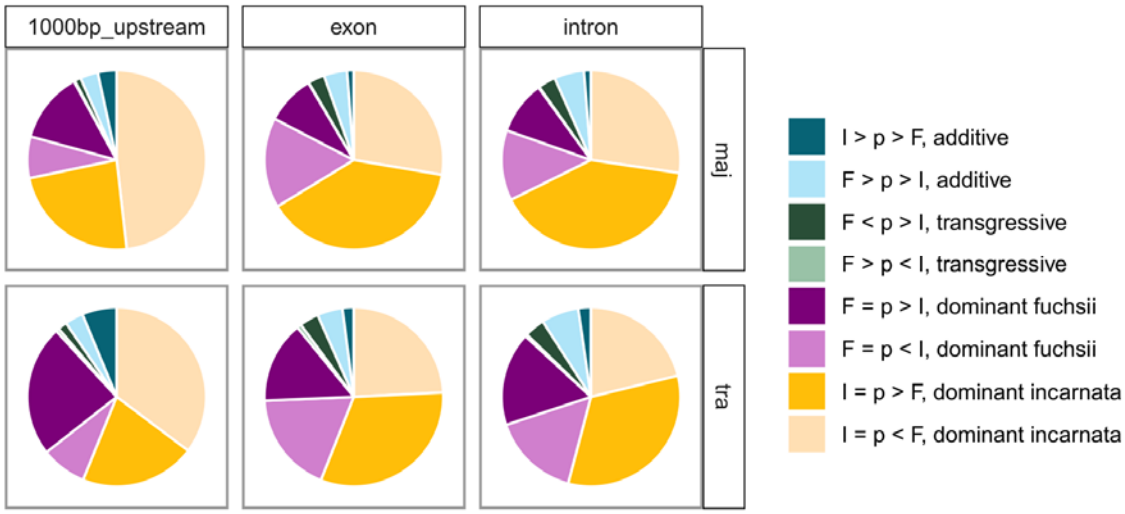

Figure S5 Proportions of peaks within each genomic interaction (additive, transgressing or dominant to either diploid) for each genomic region (1000 bp upstream/promoter, exon, intron). Colours are according to legend. Letters in legend represent **F** - *D. fuchsii*, **p** - polyploid and **I** - *D.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

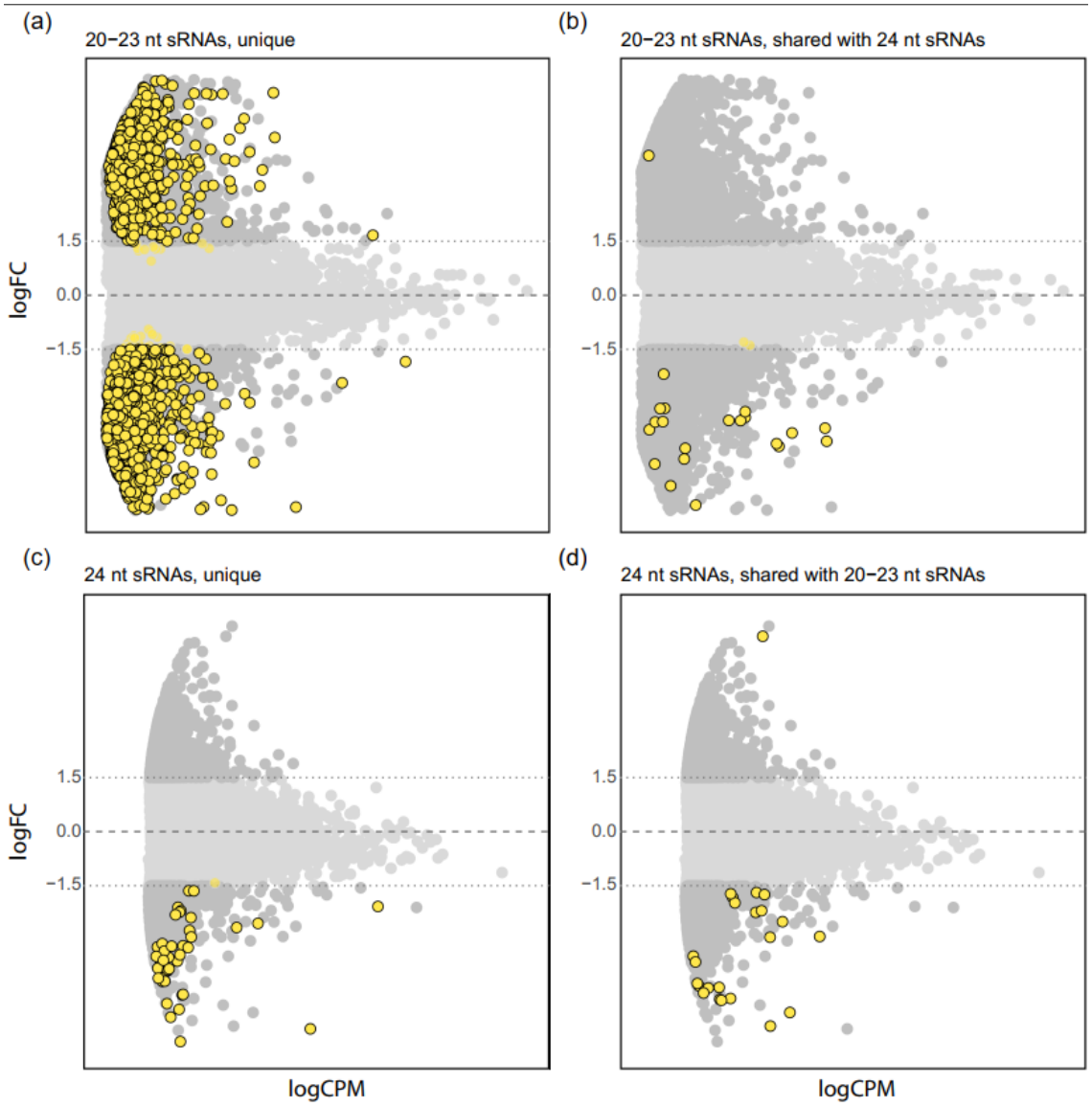

Figure S6 MA plots showing differences between *D. majalis* and *D. traunsteineri* for TEs types grouped together, the central dashed line marks 0 while the dotted lines mark logFC -1.5 and 1.5. Negative logFC values represent over-targeting in *D. majalis* as compared to *D. traunsteineri*, whereas 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.

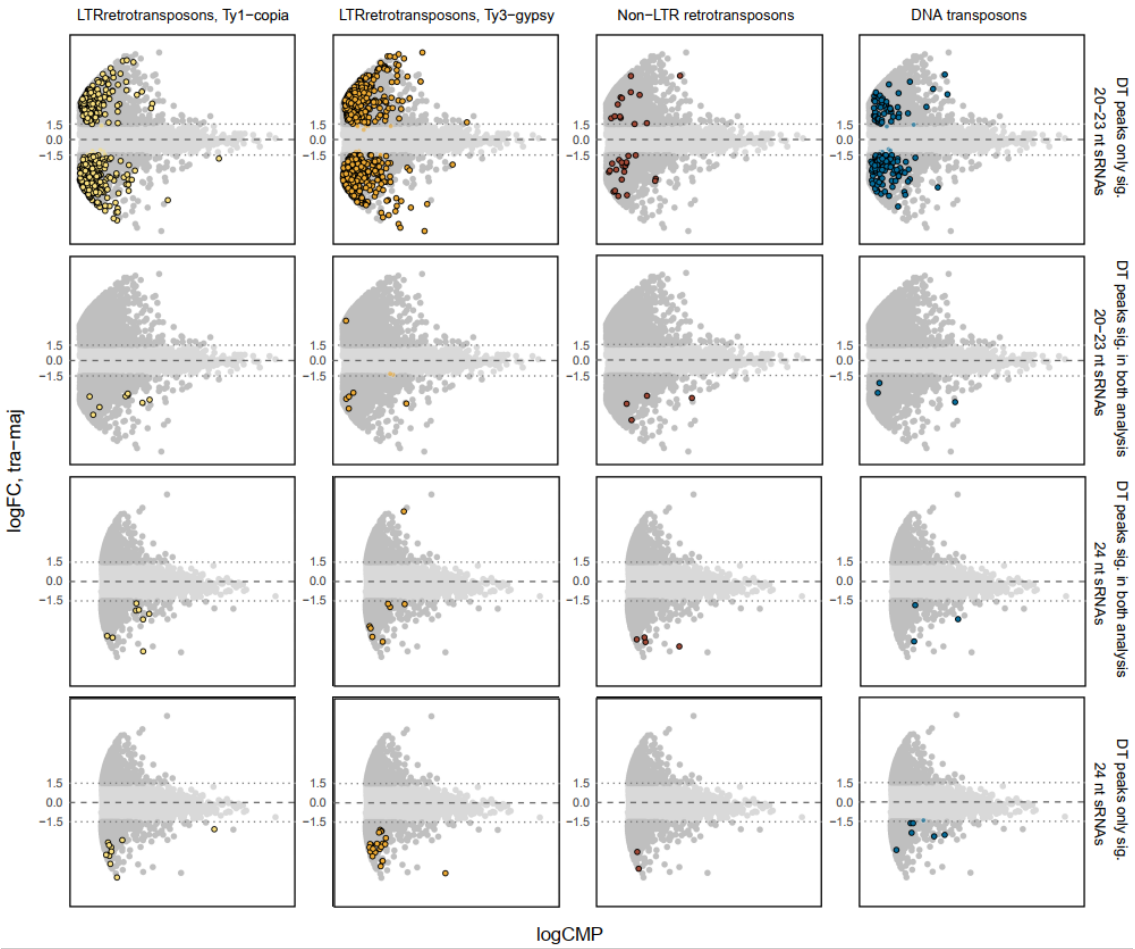

Figure S7 MA plots showing differences between *D. majalis* and *D. traunsteineri* for TEs, the central dashed line marks 0 while the dotted lines mark logFC -1.5 and 1.5. Negative logFC values represent over-targeting in *D. majalis* as compared to *D. traunsteineri*, whereas positive logFC values indicate 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.

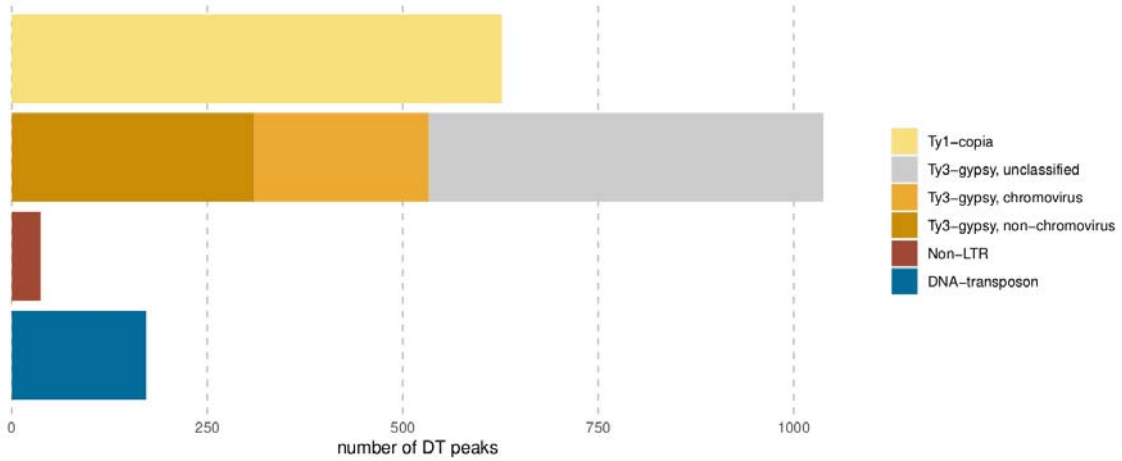

Figure S8 Count of DT peaks between *D. majalis* and *D. traunsteineri*, the coloured points of the plots in the 1st row (i.e. 20−23 nt smRNAs) in Fig. S7. Here Ty3-gypsy has been divided into the next annotation level.

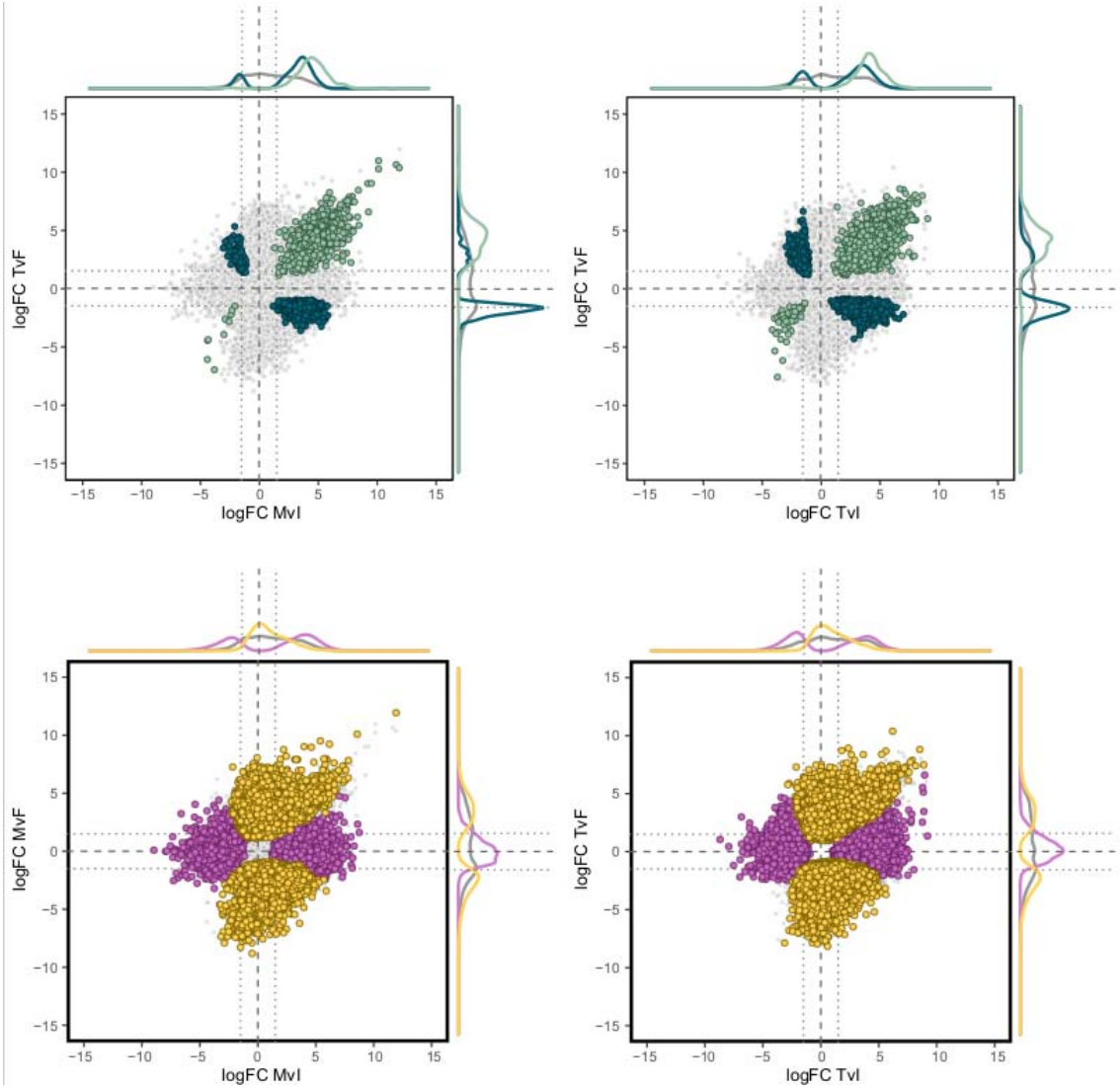

Figure S9 Patterns of genomic interactions comparing logFC values for *D. majalis* (left) and *D. traunsteineri* (right) towards either diploid, *D. fuchsii* on the Y-axis and *D. incarnata* on the X-axis for 20-23 nt smRNAs peaks found in annotated TEs. Top row shows with coloured symbols transgressive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

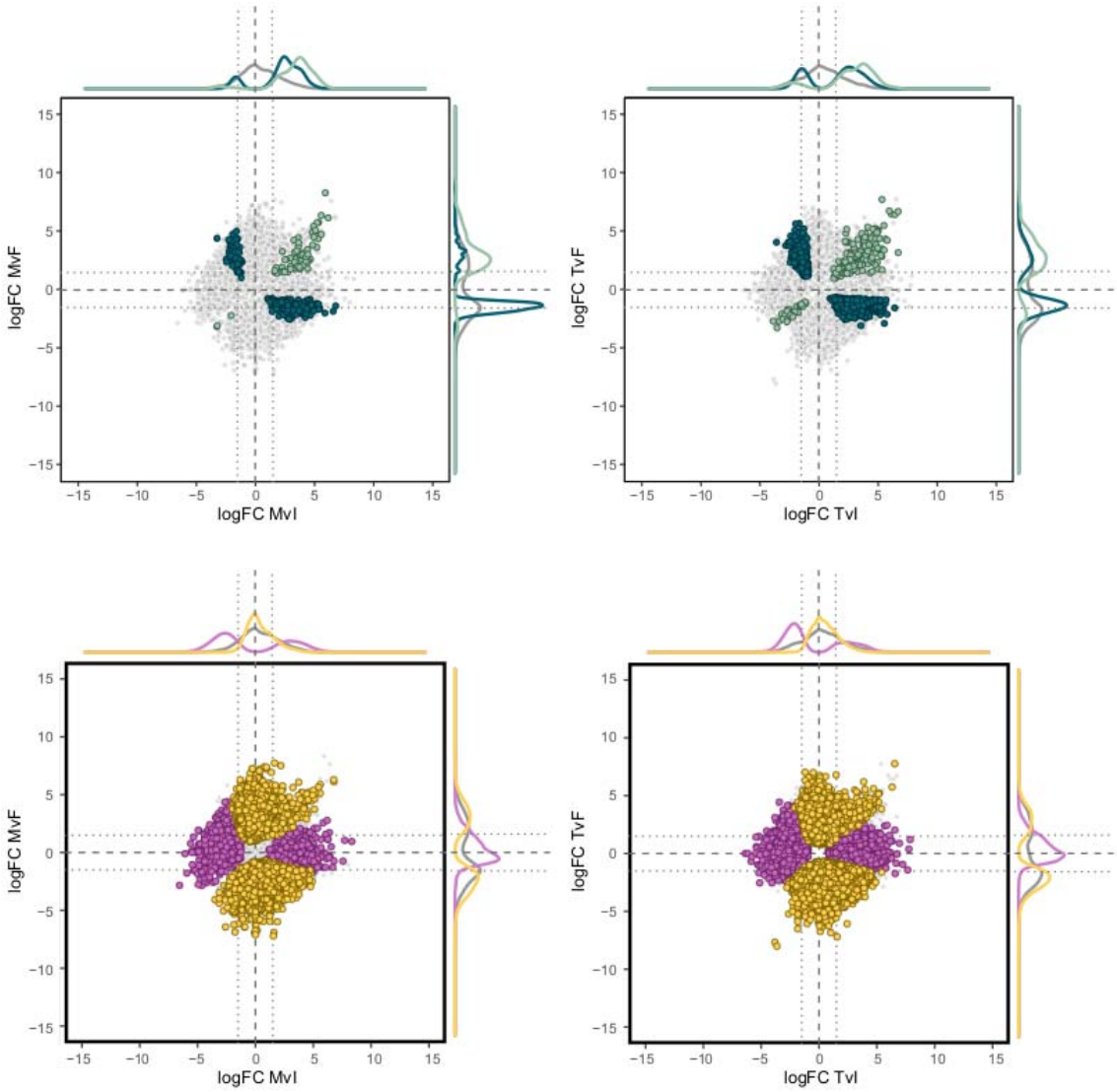

Figure S10 Patterns of genomic interactions comparing logFC values for *D. majalis* (left) and *D. traunsteineri* (right) towards either diploid, *D. fuchsii* on the Y-axis and *D. incarnata* on the X-axis for 24 nt smRNAs peaks found in annotated TEs. Top row shows with coloured symbols transgressive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